

预冷却引射系统性能一维分析*

张建强^{1,2}, 王振国^{1,2}, 李清廉², 徐万武², 邹建军²

(1. 国防科技大学 高超声速冲压发动机技术重点实验室, 湖南 长沙 410073;

2. 国防科技大学 航天科学与工程学院, 湖南 长沙 410073)

摘要:根据引射器的一维设计理论可知,二次流在进入混合室之前进行预冷降温可以提高引射效率,增大引射系数,但引入预冷器会同时引起流动损失,故需要对引射系统进行性能评估。针对设有预冷器的引射系统,应用一维理论分析预冷对系统性能的影响,重点分析预冷增强效果与流阻减弱效果对引射效率的作用。研究发现:预冷器对引射系统同时带来冷却增强作用和流阻减弱作用,横截面积和换热面积是主要影响因素。预冷器存在临界横截面积,横截面积大于临界值时,换热面积越大,引射性能越高;反之,换热面积越大,引射性能越低。等压混合引射方案比等截面混合引射方案性能高,前者引射系数比后者大60%;预冷却能够有效提高引射性能,尤其是等截面混合引射方案,性能提高可达35.5%。

关键词:一维分析;预冷增强;流阻减弱;临界横截面积

中图分类号:V435 **文献标志码:**A **文章编号:**1001-2486(2017)03-001-06

One-dimensional analysis for performance of ejector with precooling

ZHANG Jianqiang^{1,2}, WANG Zhenguo^{1,2}, LI Qinglian², XU Wanwu², ZOU Jianjun²

(1. Science and Technology on Scramjet Laboratory, National University of Defense Technology, Changsha 410073, China;

2. College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: According to the one-dimension design theory of ejector, cooling the secondary flow before it enters into the mixing chamber can promote the eject efficiency and increase the eject coefficient, but flow loss is brought with the addition of precooler, so the performance evaluation of ejector is necessary. Looking on the eject system with precooler, the effect of precooler on the performance of system was analyzed through the one-dimension theory, and the effects of the intensifying effect of precooling and the weakening effect of resistance on eject efficiency were analyzed emphatically. The research results indicate that; the precooler brings intensifying effect of precooling and weakening effect of resistance to eject system, and the cross section area and heat transfer area of the precooler are the dominating factors; cross section area of the precooler has a critical value, when the cross section area is bigger than the critical value, the eject performance improves with the increase of heat transfer area, otherwise it worsens; the eject performance of the equivalent pressure mixing scheme is better than that of the equivalent area mixing scheme, the eject coefficient of the former is 60% higher than the latter; precooling improves the eject performance effectively, especially for the equivalent area mixing scheme, the performance is improved by 35.5%.

Key words: one-dimensional analysis; intensifying effect of precooling; weakening effect of resistance; critical cross section area

目前,主动引射高空模拟试车台压力恢复系统主要有以下三种方案:“扩压器+冷却室+隔离阀+蒸汽引射器”“扩压器+环状蒸汽引射器+冷却室+排气机组+引射器”“扩压器+冷却室+排气机组”。其中,冷却室是三种方案都存在的部件。

对于引射方案来说,当模拟马赫数高时,燃气总温很高。亚-超引射方案中高超声速射流在扩压器内通过激波串的压缩作用变为亚声速气流,在压力升高的同时气流静温也恢复到接近气流总温的程度,高静温、高静压将导致扩压器和引射器结构热载荷大大增加,故需要采用主动冷却措施或者特殊的热防护结构。按照引射器一维理论,在保持引射性能不变的前提下,不论引射气流与被引射气流的比热比、分子量等物性参数是否相同,都存在 $n \propto 1/\sqrt{\theta}$ 的关系(其中 n 为引射系数,是被引射工质与引射工质质量流量之比; θ 为总温比,是被引射工质与引射工质总温之比)^[1-6]。引射器一维理论表明随着总温比的减小,引射系

* 收稿日期:2016-01-10
基金项目:新世纪优秀人才支持计划资助项目(NCET-13-0156)
作者简介:张建强(1987—),男,山东泗水人,博士研究生,E-mail:nabiandesha@163.com;
王振国(通信作者),男,教授,博士,博士生导师,E-mail:Zhenguo_Wang@nudt.edu.cn

数增大,引射工质质量流量减小,运行成本降低,并且引射器结构尺寸减小,建设成本降低。

综合来看,对工质进行较大程度地冷却,不仅可以解决热防护问题,而且可以有效地增大引射系数,减小引射工质质量流量,降低运行成本和建设成本,其必要性显而易见^[6]。下面针对自由射流系统中的引射系统进行预冷却方案设计。

1 预冷却引射系统

预冷却引射系统由预冷扩压器、预冷器和引射器组成,其中预冷扩压器对壁面进行冷却热防护,预冷器对二次流进行冷却,引射器将二次流引射排出。图 1 和图 2 分别为预冷却等截面混合及等压混合引射系统示意图。图中标号 S, HX, HX', 1, 2, 3, 4 分别表示各特征截面,参数 m_2 , P_{02} , p_2 , T_{02} , Cp_2 , M_2 , γ_2 , V_2 , Ma_2 , A_2 分别表示截面 2 上气流流量、总压、静压、总温、定压比热、分子量、比热比、速度、马赫数以及截面面积,后文中各参数沿用该表述方式。

1.1 预冷扩压器

预冷扩压器由等直段和扩张段组成,壁面加工冷却槽道用于防热。超声速气流在等直段经斜

激波串减速为亚声速,然后通过扩张段进一步减速增压。

1.2 预冷器

预冷器采用叉排管束换热器方案,图 3 为预冷器结构示意图,换热管内的冷却剂和管外主流气体对流换热。换热过程分三部分:冷却剂与内管壁的对流换热;换热管内外壁之间的导热;主流气体与外管壁的对流换热。

预冷器结构参数主要包括横截面积、换热管总面积、管径 d_0 、壁厚、横向间距 s_1 、纵向间距 s_2 ,冷却剂选择水,其工况参数主要包括流量、入口温度、压力。

1.3 引射器

引射器由二次流入口通道、引射喷管、混合室、扩压器组成,引射器扩压器由等直段和扩压段组成。一次流和二次流分别由引射喷管和二次流入口通道进入混合室,经混合后气流进入扩压器,在等直段经过斜激波串减速为亚声速,然后通过后面的扩压段进一步减速增压。一维引射理论主要公式如式(1)~(3)所示,分别为质量、动量和能量守恒公式。

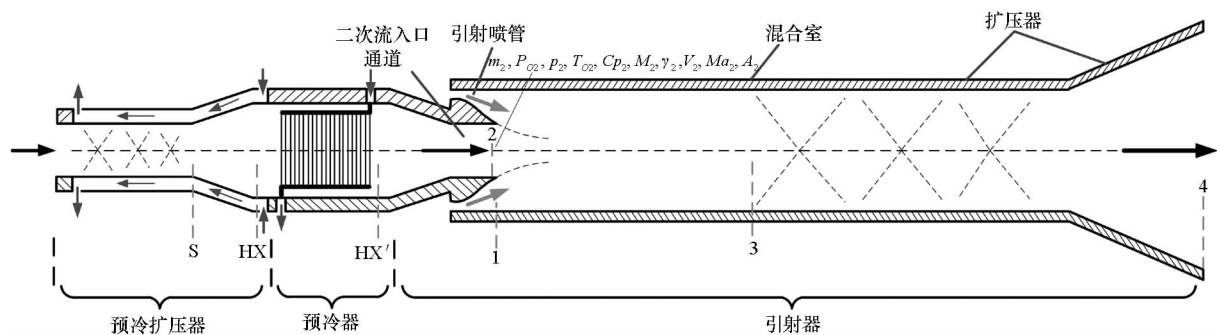


图 1 预冷却等截面混合引射方案示意图

Fig. 1 Schematic diagram of precooled equivalent area mixing scheme

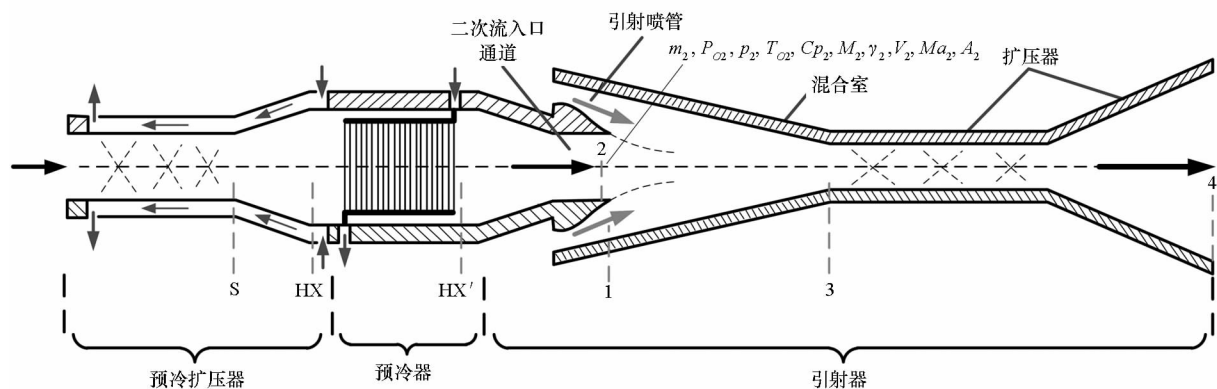
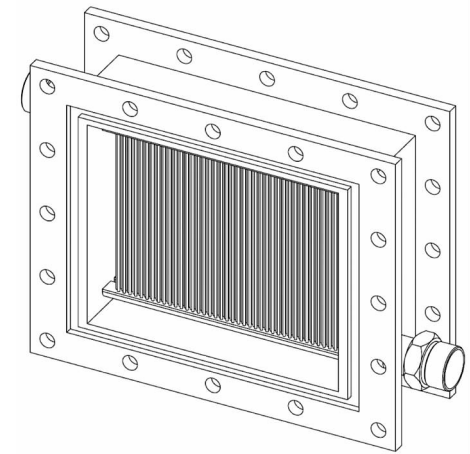


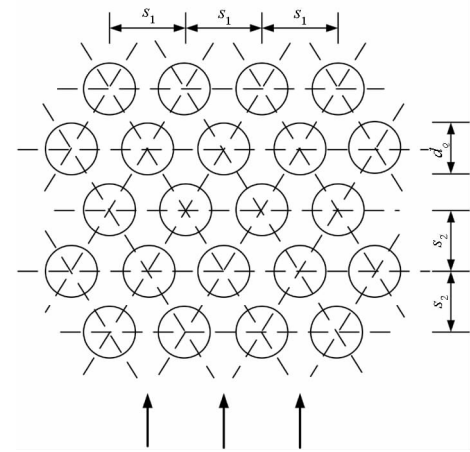
图 2 预冷却等压混合引射方案示意图

Fig. 2 Schematic diagram of precooled equivalent pressure mixing scheme



(a) 结构总图

(a) Overall view of the structure



(b) 叉排管束示意图

(b) Schematic diagram of staggered tube bundles

图3 预冷器结构示意图

Fig.3 Sketch of structure of the precooler

$$m_3 = (1 + n) m_1 \tag{1}$$

$$m_3 V_3 - (m_1 V_1 + m_2 V_2) = p_1 A_1 + p_2 A_2 - p_3 A_3 \tag{2}$$

$$m_1 C_{p1} T_{01} + m_2 C_{p2} T_{02} = m_3 C_{p3} T_{03} \tag{3}$$

其中, $n = m_2 / m_1$ 为引射系数。

2 计算流程和方法

该问题中,预冷扩压器入口经斜激波串后截面S上的气流参数为:总压 62 kPa;静压 58.4 kPa (0.3Ma);总温 1556.6 K;分子量 28.75;比热比 1.339。计算过程中改变预冷器结构参数和工况参数,研究引射性能的变化规律,引射器出口压力确定,引射系数越大,引射器性能越高。

计算过程作如下基本假设:

- 1) 气体为量热完全气体;
- 2) 气体参数沿轴线一维均匀变化,各截面参数一致;

3) 预冷扩压器和引射器扩压器中的斜激波串按照正激波处理,亚声速扩压段内气体流动按照等熵过程处理;

4) 预冷扩压器冷却槽道主要用于壁面防热,对气流的冷却作用忽略不计。

预冷器性能按照工程设计方法估算^[7-12],引射器参数计算参照一维设计理论,由三大方程联立求解混合过程,图4为预冷却引射方案参数计算流程。

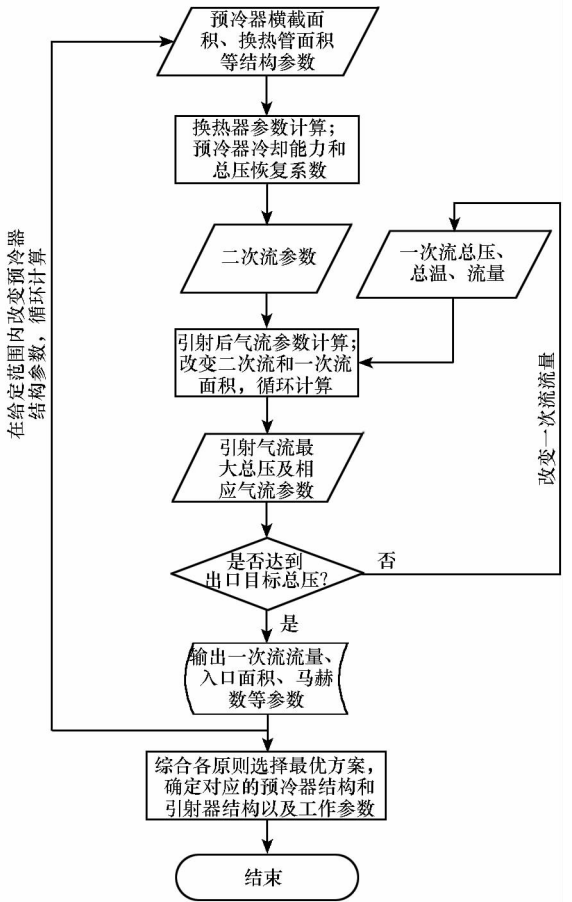


图4 预冷却引射方案参数计算流程图^[6]

Fig.4 Calculation flow chart of precooled ejector parameters^[6]

预冷却引射方案参数的具体计算过程如下:

- 1) 选定引射器结构参数和冷却剂工况参数,估算其换热功率和流动损失,求得其总温恢复系数和总压恢复系数,确定换热器出口气流参数。
- 2) 选定一次流流量、总压和总温,改变一次流和二次流面积以改变入口马赫数,循环计算求得最大出口总压,确定对应的入口面积。如果该总压达到出口压力,则停止计算,反之则改变一次流流量,直到引射器出口最大总压达到要求,确定对应的一次流流量。
- 3) 在一定范围内改变引射器结构参数和冷却剂工况参数,返回步骤1循环计算。综合考虑

建设、运行成本以及结构占地等因素,选择最优方案,确定对应的预冷器结构和引射器结构以及工作参数。

3 计算结果和分析

选择来流面积 A_s 作为基准面积,即为总温 1556.6 K、总压 62 kPa、马赫数 (Ma) 为 0.3 下的流通面积,实际面积与基准面积之比即为后文中的无量纲面积。一次流总温 1200 K、总压 3 MPa 保持不变。为了将二次流顺利引射排入大气,对出口目标总压留取 50% 的余量,即为 1.5 bar,二次流增压比为 2.42。

3.1 预冷器

计算中保持冷却剂工况参数不变,改变预冷器结构参数,研究预冷器性能的变化规律。冷却水无量纲流量(冷却水流量/二次流流量)为 1.294,入口温度为 300 K、压力为 5 MPa。

预冷器设计点结构参数分别为:无量纲横截面积 $A_{HX} = 4.23$,换热管总面积 $A_{tube} = 44.06$,换热管外径 0.98 mm,壁厚 0.04 mm,相对横向间距 S_1 ,相对纵向间距 S_2 。性能参数包括总压恢复系数 Po_{rec} 和总温恢复系数 To_{rec} ,分别为气流在预冷器出口与进口的总压和总温之比,表征气流经过预冷器后的流动损失和换热功率。 Po_{rec} 越大, To_{rec} 越小,说明预冷器流动损失越小,功率越高,性能越优。预冷器设计点性能参数分别为 $Po_{rec} = 0.94$, $To_{rec} = 0.515$ 。

计算中依次改变各个结构参数,保持其他参数不变,研究单一变量的影响,实际参数与设计点参数之比为相对参数。由计算结果可知,预冷器横截面积和换热面积是影响其性能的主要因素,如图 5 和图 6 所示,其中 A'_{HX} 和 A'_{tube} 分别为相对横截面积和相对换热面积。 A'_{HX} 由 0.5 增大至 1.25, Po_{rec} 增大 67.8%, To_{rec} 增大 37.5%, A'_{tube} 由 0.4 增大至 1.6, Po_{rec} 减小 8.7%, To_{rec} 减小 47.3%。随着 A'_{HX} 的增大,总压恢复系数不断增大但增幅逐渐减小,总温恢复系数则基本线性增大;随着 A'_{tube} 的增大,预冷器总温和总压恢复系数均基本呈线性减小。

随着横截面积的增大,气流速度减小,流动损失相应减小,但气流与换热管对流换热系数也随之减小,使得换热功率降低。换热管总面积增大,换热功率相应增大,预冷器出口气流温度降低,但换热管管排数随之增大,流动损失增大。

3.2 预冷却引射方案

常规的非预冷亚超引射方案结果如图 7 和

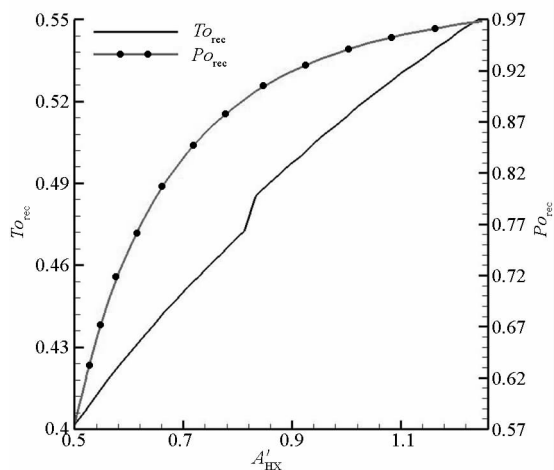


图 5 预冷器性能随横截面积变化曲线

Fig. 5 Evolution curve of precooler performance along with cross section area

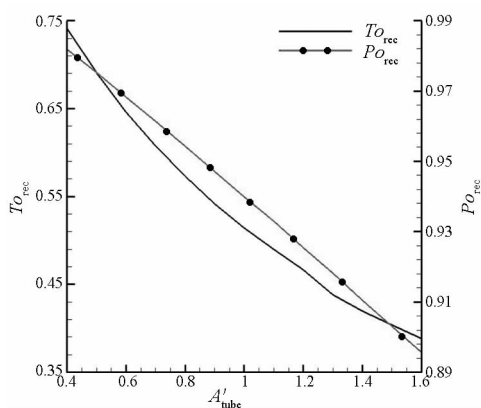


图 6 预冷器性能随换热面积变化曲线

Fig. 6 Evolution curve of percooler performance along with heat transfer area

图 8 所示,两图分别为等截面混合引射方案和等压混合引射方案。当引射系数确定时,通过调整二次流和主流入口面积,可以使出口气流总压达到最大,从而使引射性能达到最优。

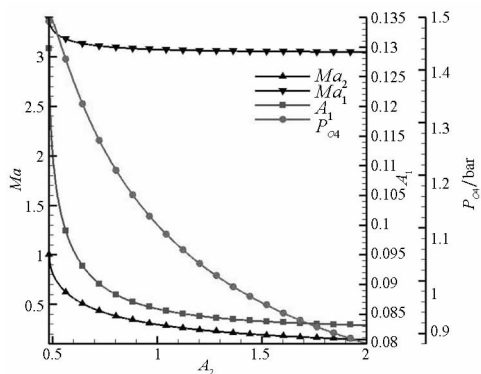


图 7 等截面混合引射方案参数变化曲线

Fig. 7 Evolution curves of ejector parameters of equivalent area mixing scheme

等截面混合引射方案下,随着 A_2 的减小, Ma_2 不断增大,静压降低, Ma_1 和 A_1 相应地增大, P_{04} 不断增大。当引射系数 $n=0.864$ 时, $A_2=0.489$ 对应的出口总压达到最大,为 1.5 bar,达到引射要求,相应的 $A_1=0.12$, $Ma_1=3.32$, $Ma_2=0.89$,此即非预冷等截面混合引射方案的最佳工况。

等压混合引射方案下,随着 A_2 的减小, Ma_2 不断增大,静压降低, Ma_1 和 A_1 相应地增大, P_{04} 先增大后减小。当引射系数 $n=1.382$ 时, $A_2=0.505$ 对应的出口总压达到最大,为 1.5 bar,达到引射要求,相应的 $A_1=0.069$, $Ma_1=3.26$, $Ma_2=0.79$,混合室收缩比为 0.716,此即非预冷等压混合引射方案的最佳工况。

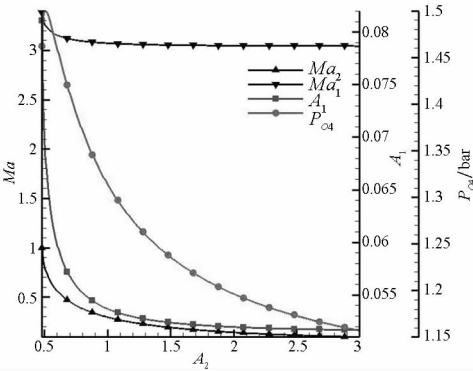


图8 等压混合引射方案参数变化曲线
Fig. 8 Evolution curves of ejector parameters of equivalent pressure mixing scheme

由上述计算可知,等压混合引射方案比等截面混合引射方案性能高,同等增压比前提下,前者引射系数比后者大约 60%,同时系统尺寸较小。

增加预冷器会使二次流温度降低,冷却作用能够提高引射性能,但同时也会引起流动损失,而流阻作用又会降低引射性能。以下针对两种预冷引射方案进行性能计算,主要考察引射系数随预冷器换热面积和横截面积的变化曲线,相对换热面积 A'_{tube} 取 0.4 ~ 1.6,相对横截面积 A'_{HX} 取 0.5 ~ 1.25,计算结果反映了预冷器对引射性能的影响。

3.2.1 预冷却等截面混合引射方案

图9为预冷等截面混合引射方案引射系数变化曲线,由结果可知,保持 A'_{tube} 不变,随着 A'_{HX} 的增大,流动损失减小,冷却作用同时减弱, n 逐渐增大至趋于平稳,可见流阻减弱作用大于冷却增强作用,但两者之间的差距逐渐减小。保持 A'_{HX} 不变,随着 A'_{tube} 的增大,流动损失增大,冷却作用同时增强, n 则呈现出相反的变化规律:临界相对横截面积约为 0.8,当 $A'_{HX}=0.8$ 时, n 随着 A'_{tube} 的改变基本保持不变,即换热面积对引射性能几乎

没有影响;当 $A'_{HX} > 0.8$ 时, n 随着 A'_{tube} 的增大而增大,换热面积增加带来的冷却增强作用的增长速度大于流阻减弱作用的增长速度,并且两者之间的差别随着横截面积的增大而增大;当 $A'_{HX} < 0.8$ 时, n 随着 A'_{tube} 的增大而减小,换热面积增加带来的冷却增强作用的增长速度小于流阻减弱作用的增长速度,并且两者之间的差别随着横截面积的减小而增大。

非预冷等截面混合引射方案 $n=0.864$,对应图9中的虚线,虚线以下表示流阻减弱作用大于冷却增强作用,虚线以上表示流阻减弱作用小于冷却增强作用。预冷器设计参数应该在虚线以上,才能体现预冷对引射系统的增强效果。当横截面积大于临界值时, A'_{tube} 越大,引射性能增强效果越明显,当 $A'_{HX}=1.25$ 、 $A'_{tube}=1.6$ 时,引射系数 $n=1.171$,此时引射系统的性能最优,比非预冷引射方案的提高 35.5%。

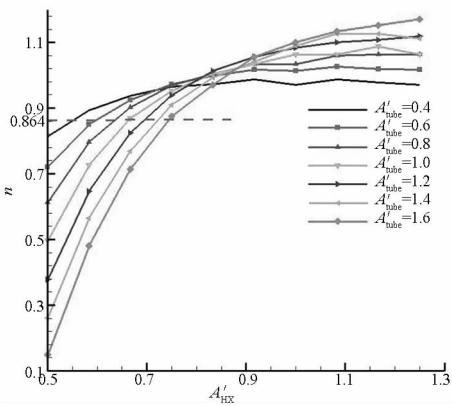


图9 预冷却等截面混合引射方案引射系数变化曲线
Fig. 9 Evolution curve of eject coefficient of precooled equivalent area mixing scheme

3.2.2 预冷却等压混合引射方案

图10为预冷却等压混合引射方案引射系数变化曲线,由结果可知,保持 A'_{tube} 不变,随着 A'_{tube} 的增大,流动损失减小,冷却作用同时减弱, n 逐渐增大至趋于平稳,可见流阻减弱作用大于冷却增强作用,但两者之间的差距逐渐减小。保持 A'_{HX} 不变,随着 A'_{tube} 的增大,流动损失增大,冷却作用同时增强, n 同样呈现出相反的变化规律:临界相对横截面积约为 1.1,当 $A'_{HX}=1.1$ 时, n 随着 A'_{tube} 的改变基本保持不变,即换热面积对引射性能几乎没有影响;当 $A'_{HX} > 1.1$ 时, n 随着 A'_{tube} 的增大而增大,换热面积增加带来的冷却增强作用的增长速度大于流阻减弱作用的增长速度,并且两者之间的差别随着横截面积的增大变化较小;当 $A'_{HX} < 1.1$ 时, n 随着 A'_{tube} 的增大而减小,换热面积增加带来的冷却增强作用的增长速度小于流阻减弱作

用的增长速度,并且两者之间的差别随着横截面积的减小而增大。

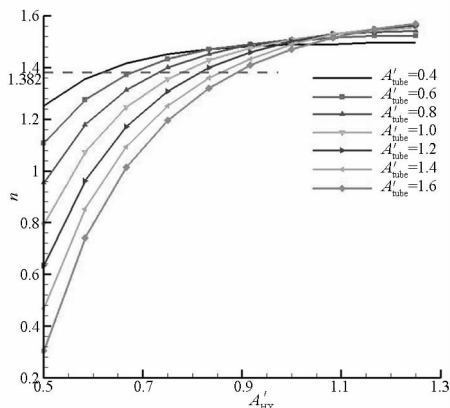


图 10 预冷却等压混合引射方案引射系数变化曲线

Fig. 10 Evolution curve of eject coefficient of precooled equivalent pressure mixing scheme

非预冷等压混合引射方案 $n = 1.382$, 对应图 10 中的虚线, 虚线以下表示流阻减弱作用大于冷却增强作用, 虚线以上表示流阻减弱作用小于冷却增强作用。预冷器设计参数应该在虚线以上, 才能体现预冷对引射系统的增强效果。当横截面积大于临界值时, A'_{tube} 越大, 引射性能增强效果越明显, $A'_{\text{HX}} = 1.25$ 、 $A'_{\text{tube}} = 1.6$ 时, 引射系数 $n = 1.57$, 此时引射系统的性能最优, 比非预冷引射方案的提高 13.6%。

由计算结果可知, 预冷却引射方案能够有效增强引射性能, 尤其是对于等截面混合引射方案, 性能提高达 35.5%。预冷器对引射系统同时具有冷却增强作用和流阻减弱作用, 必须保证设计点处于图 9 和图 10 虚线以上区域才能体现预冷过程的优越性; 同时应根据横截面积选择换热面积, 当横截面积大于临界值时, 换热面积越大, 引射性能越高; 反之, 换热面积越小, 引射性能越高。

4 结论

针对预冷却引射系统开展性能分析计算, 利用引射一维理论和换热器工程设计方法建模, 得到预冷却对引射性能的影响, 并发现了预冷器的临界横截面积, 所得结果能够有效指导系统设计。主要结论如下:

1) 预冷器对引射系统同时带来冷却增强作用和流阻减弱作用, 横截面积和换热面积是主要影响因素, 两者在增强冷却和减弱流阻上起相反的作用;

2) 预冷器存在临界横截面积, 横截面积大于临界值时, 换热面积越大, 引射性能越高, 冷却带来的

引射增强作用起主导作用; 反之, 换热面积越大, 引射性能越低, 流阻带来的引射减弱作用起主导作用;

3) 等压混合引射方案比等截面混合引射方案性能高, 前者引射系数比后者大 60%, 预冷却能够有效提高引射性能, 等截面混合引射方案性能提高可达 35.5%, 等压混合方案性能提高 13.6%。

参考文献 (References)

- [1] Nagaraja K S. Some ejector characteristics [C]//Proceedings of Aircraft Systems and Technology Conference, AIAA 1981 - 1679, 1981.
- [2] Keenan J H, Neumann E P, Lustwerk F. An investigation of ejector design by analysis and experiment [J]. Journal of Applied Mechanics, 1950, 17(3): 299 - 309.
- [3] Keenan J H, Neumann E P. A simple air ejector [J]. Journal of Applied Mechanics—Transactions of the ASME, 1942, 64: A75 - A81.
- [4] 邹建军, 周进, 徐万武, 等. 超声速环形引射器空气引射启动特性试验 [J]. 国防科技大学学报, 2008, 30(1): 1 - 4. ZOU Jianjun, ZHOU Jin, XU Wanwu, et al. Experimental investigation on the start performances of the supersonic annular air ejector [J]. Journal of National University of Defense Technology, 2008, 30(1): 1 - 4. (in Chinese)
- [5] 徐万武. 高性能、大压缩比化学激光器压力恢复系统研究 [D]. 长沙: 国防科技大学, 2003. XU Wanwu. Study of high performance, high compression ratio pressure recovery system for chemical laser [D]. Changsha: National University of Defense Technology, 2003. (in Chinese)
- [6] 吴庆伟. 预冷器及预冷引射系统性能研究 [D]. 长沙: 国防科技大学, 2014. WU Qingwei. Research on precooler and characteristics of precooling ejector system [D]. Changsha: National University of Defense Technology, 2014. (in Chinese)
- [7] Shah R K, Mueller A C, Sekulic D P. Heat exchanger [M]. Weinheim, Germany: Ullmann's Encyclopedia of Industrial Chemistry, 1988.
- [8] 杨世铭, 陶文铨. 传热学 [M]. 4 版. 北京: 高等教育出版社, 2006. YANG Shiming, TAO Wenquan. Heat transfer [M]. 4th ed. Beijing: Higher Education Press, 2006. (in Chinese)
- [9] 余建祖. 换热器原理与设计 [M]. 北京: 北京航空航天大学出版社, 2006. YU Jianzu. Principle and design of heat exchanger [M]. Beijing: Beihang University Press, 2006. (in Chinese)
- [10] 马小明, 钱颂文, 朱东生, 等. 管壳式换热器 [M]. 北京: 中国石化出版社, 2010. MA Xiaoming, QIAN Songwen, ZHU Dongsheng, et al. Shell and tube heat exchanger [M]. Beijing: China Petrochemical Press, 2010. (in Chinese)
- [11] 沙拉. 塞库利克. 换热器设计技术 [M]. 程林, 译. 北京: 机械工业出版社, 2010. Shah R K, Sekuli D P. Fundamentals of heat exchanger design [M]. Translated by CHENG Lin. Beijing: China Machine Press, 2010. (in Chinese)
- [12] 兰州石油机械研究所. 换热器 [M]. 2 版. 北京: 中国石化出版社, 2013. Lanzhou Petroleum Machinery Research Institute. Heat exchanger [M]. 2nd ed. Beijing: China Petrochemical Press, 2013. (in Chinese)

重力梯度对超大柔性空间结构在轨动力学特性的影响*

穆瑞楠¹, 谭述君², 吴志刚^{1,2}

(1. 大连理工大学 工业装备结构分析国家重点实验室, 辽宁 大连 116024;

2. 大连理工大学 航空航天学院, 辽宁 大连 116024)

摘要:空间太阳能电站是一种具有超大和高柔性特征的空间结构,这种空间结构在尺寸上远超以往的航天器,给轨道动力学特性的研究带来了新现象与新问题。以千米量级的哑铃模型为研究对象,考虑重力梯度影响,建立了 Hamilton 体系下的在轨动力学模型,利用辛龙格库塔法得到了不同参数取值下的动力学响应。通过对比仿真结果,得到了结构尺寸与重力梯度对轨道运动、姿态运动影响的定量关系。仿真结果表明,重力梯度引起了姿态-柔性振动耦合现象,姿态运动影响了结构振动曲线外部包络线样式,而柔性振动改变了姿态运动周期。

关键词:空间太阳能电站;超大柔性空间结构;Hamilton 方程;哑铃模型;重力梯度;耦合

中图分类号:V11 **文献标志码:**A **文章编号:**1001-2486(2017)03-007-08

Effect of gravity gradient on dynamical characteristics of very large flexible space structures in orbit

MU Ruinan¹, TAN Shujun², WU Zhigang^{1,2}

(1. State Key Laboratory of Structural Analysis for Industrial Equipment, Dalian University of Technology, Dalian 116024, China;

2. School of Aeronautics and Astronautics, Dalian University of Technology, Dalian 116024, China)

Abstract: Space solar power station is a kind of space structure with large size and high flexibility. It is far larger than the previous spacecraft in size, which results in new phenomena and new problems on the study of dynamical characteristics. The kilometer-scale dumbbell model was studied. The Hamilton's dynamical model on orbit was established under the effect of gravity gradient. The symplectic Runge-Kutta method was used with different combinations of parametrical values to obtain dynamical responses. By comparing the simulation results, the quantitative relationships were determined respectively between the size of space structure and the effect of gravity gradient on orbital motion and attitude motion. It is found that: due to the gravity gradient, the coupling phenomenon between attitude motion and elastic vibration occurs; the attitude motion has great influence on the external envelope curve of elastic vibration response, while the period of it is changed by elastic vibration.

Key words: space solar power station; very large flexible space structure; Hamilton equation; dumbbell model; gravity gradient; coupling

早在 1968 年,美国科学家 Glaser 就首先提出了建造空间太阳能电站的构想^[1]。因其可实现连续工作、能量利用率高等诸多优点,受到美国、日本、欧洲国家等发达国家的重点关注,并相继开展了大量的研究工作^[2-5]。目前,国际上提出的空间太阳能电站的概念设计已达几十种。美国 NASA 先后提出了“1979SPS 基准系统”以及“集成对称聚光系统”^[5],日本 JAXA 也先后提出了“SPS2003 系统”以及“分布式绳系太阳能电站卫星”^[6-7],欧空局提出了“太阳帆塔”^[8]。在 NASA 创新概念项目支持下,由美国、日本和英国科学家于 2012 年共同提出了一种新的空间太阳能电站概念方案“SPS-ALPHA”^[8]。

在众多的概念设计中有一个共同点,那就是结构尺寸都达到千米量级,远远超过目前低地球轨道上最大的航天器——国际空间站,具有超大和高柔性的结构特性。这种超大和高柔性的空间结构给动力学特性的研究带来了姿态-柔性振动耦合等复杂的新现象和新问题。不论是在轨组装,还是长时间在轨运行,都需要对结构的轨道和姿态响应做出准确的预测。因此,在轨动力学特性分析是影响这种超大柔性空间结构研究和发展的的重要因素。已有很多学者利用简单的哑铃结构开展了大型空间结构的动力学特性与耦合关系的

* 收稿日期:2016-02-02
基金项目:国家自然科学基金资助项目(11572069,11372056,11432010,11502040)
作者简介:穆瑞楠(1990—),男,辽宁大连人,博士研究生,E-mail:mrn2013@mail.dlut.edu.cn;
吴志刚(通信作者),男,教授,博士,博士生导师,E-mail:wuzhg@dlut.edu.cn

研究。Malla^[9]针对一维哑铃模型提出了考虑重力梯度的 Lagrange 形式的动力学方程,并通过比较在不同初始条件、质量比、轨道高度、轨道偏心率下的数值仿真结果,研究了结构轴向变形、姿态运动和轨道运动之间的规律,发现了三者之间复杂的耦合关系,同时研究了热辐射对耦合关系的影响。Ishimura^[10]基于 Malla^[9]建立的动力学方程,将日本的绳系空间太阳能电站结构简化为哑铃模型,并在平衡位置对方程进行线性化,同样采用数值仿真的方式,研究了质量比、频率比以及长度比三个系统参数对轴向振动、姿态以及轨道之间耦合关系的影响;用线性方程的特征值来表征系统参数的影响程度,发现对轴向振动频率和轨道频率的比值影响最大,对两端集中质量的比值影响较大。Sanyal 等^[11-12]在哑铃模型的平衡位置处导出了线性化方程,研究了线性化方程在哑铃模型轴向和横向驱动力作用下的可控性问题,指出几种欠驱动可以完成对轨道、姿态、变形的控制。结合轨道角动量守恒定律,进一步采用 Routh 简化得到的降阶方程,该方程显示只利用姿态和变形驱动,就可以实现对轨道、姿态和变形的控制。以上工作均没有开展有关重力梯度对耦合关系影响的研究,而这种影响在尺寸相对较小的结构动力学中已有很多研究。Ashley^[13]引入机翼研究的模态分析方法,建立了杆和梁在小变形、大变形时的模型,研究了重力梯度对姿态和结构变形的影响;此外,还研究了旋转对姿态和结构变形的影响,发现旋转的影响与重力梯度的影响量级相同。Sincarsin 等^[14]推导出了在中心引力场的刚体所受的重力梯度力矩,并得到保留到四阶的 Taylor 展开表达式,提出了一种惯性矩的新定义,讨论了重力梯度力矩高阶项的影响,发现高阶项的作用不能被忽略。通过以上工作可以发现,重力梯度对耦合关系的影响极其重要,尤其在超大柔性空间结构在轨动力学特性与耦合关系的研究中需要被重点考虑。

1 哑铃模型的动力学方程

一维哑铃模型将结构总质量简化为两端集中质量,中间由只有轴向变形能力的柔性杆连接,哑铃模型受理想地球(质点)的重力场作用,在轨道平面内运动,如图 1 所示。尽管哑铃模型结构十分简单,但包含轨道、姿态和结构变形的耦合,适合初步的研究工作。

基于图 1 所示哑铃模型根据 Lagrange 变分原

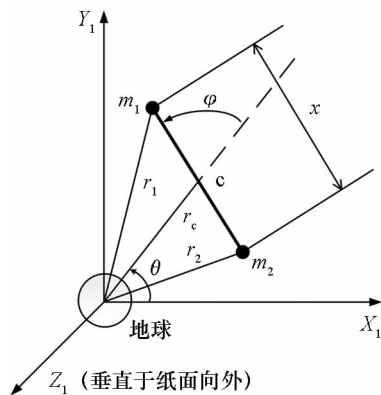


图 1 一维哑铃模型示意图

Fig. 1 Model of one-dimensional dumbbell

理,建立考虑重力梯度柔性体(Flexible Dumbbell with Gravity Gradient, FDwithGG)模型的 Hamilton 正则方程如下。

定义广义坐标为 $(r_c, \theta, \varphi, x)$ 。其中: r_c 为轨道半径,表示结构质心到地心的距离; θ 为轨道角,表示以零时刻结构质心和地心连线与当前时刻质心和地心连线之间的夹角; φ 为姿态角,表示结构轴线方向与轨道法线方向之间的夹角; x 为结构尺寸,表示两个集中质量之间的距离。则其系统动能为:

$$T = \frac{1}{2} m_c [\dot{r}_c^2 + (r_c \dot{\theta})^2] + \frac{1}{2} \bar{m} x^2 (\dot{\theta} + \dot{\varphi})^2 + \frac{1}{2} \bar{m} \dot{x}^2 \quad (1)$$

其中: m_1 和 m_2 分别为两端集中质量, $m_c = m_1 + m_2$, $\bar{m} = (m_1 \cdot m_2) / m_c$ 。可见系统动能是由轨道动能、姿态动能以及结构动能组成。系统势能为:

$$V = -\mu \left(\frac{m_1}{r_1} + \frac{m_2}{r_2} \right) + \frac{1}{2} k (x - x_s)^2 \quad (2)$$

其中: r_1, r_2 分别为两端集中质量的轨道半径; x_1, x_2 分别为两端集中质量到质心的距离; μ 为地球引力常数; $k = EA/x_s$ 为结构等效刚度系数, E 为杆的轴向弹性模量, A 为杆的截面面积; x_s 为结构原始尺寸。

定义 $(p_r, p_\theta, p_\varphi, p_x)$ 分别为广义坐标 $(r_c, \theta, \varphi, x)$ 对应的广义动量,则 Hamilton 正则方程为:

$$\begin{cases} \dot{r}_c = \frac{p_r}{m_c} \\ \dot{\theta} = \frac{p_\theta - p_\varphi}{m_c r_c^2} \\ \dot{\varphi} = \frac{p_\varphi}{m_c x^2} - \dot{\theta} \\ \dot{x} = \frac{p_x}{\bar{m}} \end{cases} \quad (3)$$

$$\begin{cases} \dot{p}_r = \frac{(p_\theta - p_\varphi)^2}{m_c r_c^3} - \mu \left[\frac{m_1(r_c + x_1 \cos \varphi)}{r_1^3} + \frac{m_2(r_c - x_2 \cos \varphi)}{r_2^3} \right] + Q_r \\ \dot{p}_\theta = Q_\theta \\ \dot{p}_\varphi = -\bar{m} \mu r_c x \sin \varphi \left(-\frac{1}{r_1^3} + \frac{1}{r_2^3} \right) + Q_\varphi \\ \dot{p}_x = \frac{p_\varphi^2}{\bar{m} x^3} - \bar{m} \mu \left[\frac{x_1 + r_c \cos \varphi}{r_1^3} + \frac{x_2 - r_c \cos \varphi}{r_2^3} \right] - k(x - x_s) + Q_x \end{cases} \quad (4)$$

其中: $Q_r, Q_\theta, Q_\varphi, Q_x$ 分别是广义坐标 $(r_c, \theta, \varphi, x)$ 对应的外部非保守广义力。式(3)和式(4)中含有地球引力常数 μ 的项为重力项。基于上面建立的 FDwithGG 模型, 通过对该模型的简化可以给出其他三种模型。如令 $x_1 = x_2 = 0$ 且 $r_1 = r_2 = r_c$, 则式(3) ~ (4) 简化为不考虑重力梯度的柔性体哑铃(Flexible Dumbbell without Gravity Gradient, FDwithoutGG)模型; 如令 $x = x_s$, 则式(3) ~ (4) 简化为考虑重力梯度的刚体哑铃(Rigid Dumbbell with Gravity Gradient, RDwithGG)模型; 如令 $x_1 = x_2 = 0, x = x_s$ 且 $r_1 = r_2 = r_c$, 不考虑柔性, 则式(3) ~ (4) 简化为不考虑重力梯度的刚体哑铃(Rigid Dumbbell without Gravity Gradient, RDwithoutGG)模型。

对比 RDwithGG 模型与 RDwithoutGG 模型, 可以研究重力梯度的影响; 而对比 FDwithGG 模型和 RDwithGG 模型可以研究在重力梯度作用下柔性振动的影响。对比四种模型中的轨道运动方程, 可以发现考虑重力梯度时重力项与姿态角有关, 而不考虑重力梯度时则无关; 不考虑重力梯度时, 姿态运动方程中不含有重力项, 只受轨道角速度以及结构变形的影响, 而考虑重力梯度时姿态运动受重力梯度、轨道角速度以及结构变形的共同影响; 柔性振动方程中存在耦合项 p_φ , 当不考虑重力梯度且外部广义力 Q_φ 为零时, p_φ 为常值, 且可以由初始条件确定, 耦合项相当于常力作用在结构上, 即结构振动只受初始条件影响。

2 模型参数的确定

考虑到“低轨组装和高轨运行”是目前空间太阳能电站的一种主要设计方案, 同时在低轨上重力梯度的影响更为显著, 所以分析中的轨道均选择轨道高度为 200 km 的低地球轨道。本文重点研究重力梯度对超大柔性空间结构动力学特性的影响, 忽略了热辐射、太阳光压等空间环境干扰。因此, 在上节建立的哑铃模型中, 广义动量导数方程右端的外部非保守力均为零(即 $Q_r = Q_\theta = Q_\varphi = Q_x = 0$)。轨道形状分为圆轨道和小偏心率

轨道($e = 0.0785$)。

2.1 物理参数及初始条件选取

哑铃模型两端集中质量为 $m_1 = m_2 = 5.0 \times 10^5 \text{ N} \cdot \text{min}^2/\text{km}$ (即 $1.8 \times 10^6 \text{ kg}$), 中间柔性杆的原始尺寸 x_s 和轴向刚度 k 分别是影响重力梯度以及柔性振动的重要参数, 不同情况对应不同取值, 若无说明则 $x_s = 1.0 \text{ km}$, $k = k_s = 1.63859 \times 10^8 \text{ N/km}$ 。涉及的常数有: 地球平均半径 $R = 6378 \text{ km}$, 万有引力常数 $G = 8.64432 \times 10^{-9} \text{ km}^4/(\text{N} \cdot \text{min}^4)$, 地球引力常数 $\mu = 1.43496 \times 10^9 \text{ km}^3/\text{min}^2$ 。

柔性体模型需要 8 个初始条件, 而刚体模型需要 6 个初始条件, 分别对应于广义坐标以及它们导数的初始值, 表示为 $r_c(0), \theta(0), \varphi(0), x(0), \dot{r}_c(0), \dot{\theta}(0), \dot{\varphi}(0)$ 以及 $\dot{x}(0)$ 。为了计算结果的有效性及其可比较性, 给出一些关于初始条件的假设。假设结构从近地点或远地点开始运动, 则 $\dot{r}_c(0) = \dot{\theta}(0) = 0, r_c(0) = 6578 \text{ km}$; 在圆轨道时, $\dot{\theta}(0) = 0.071003391567 \text{ rad/min}$, 在小偏心率轨道时, $\dot{\theta}(0) = 0.073737628934223 \text{ rad/min}$; 而 $\dot{\varphi}(0)$ 和 $\dot{x}(0)$ 没有特殊要求, 在所有情况下均假设为 0; $\varphi(0)$ 和 $x(0)$ 分别是影响重力梯度以及柔性振动的重要参数, 不同情况对应不同取值, 若无说明则 $\varphi(0) = 0, x(0) = x_s$ 。

2.2 数值求解方法的确定

哑铃模型的动力学方程通常采用的是 Lagrange 方程的形式, 而数值算法采用的是传统非保辛的龙格库塔(Runge-Kutta, RK)法^[9-12]。本文采用 Hamilton 形式的动力学模型以及文献[15]中的辛龙格库塔(Symplectic Runge-Kutta, SRK)法的数值算法。传统的 RK 法是经典的非保辛方法, 而文献[15]中的 SRK 法是经典的保辛方法。相比于 RK 法, SRK 法具有保辛优势, 能够使动量及能量两种守恒量残差不会随积分时间增大而增大, 使仿真结果更准确地反映守恒系统的本质特征, 这对于揭示长时间运行系统的特性是非常重要的。同时, Lagrange 和 Hamilton 两种模型形式在物理本质上完全等价, 但利用 Hamilton 形式动力学方程可以得到更好的系统守恒特征。设质点的初始条件为 $r_c(0) = 6578 \text{ km}$, $\dot{\theta}(0) = 0.04 \text{ rad/min}$, $\dot{r}_c(0) = \dot{\theta}(0) = 0$; 求解步长为 0.5 min, 求解时长为 1000 min。表 1 给出了 RK 法与 Hamilton 形式方程, SRK 法与 Hamilton 形式方程, 以及 SRK 法与 Lagrange 形式方程的结果。在这

个算例中 RK 法给出的残差随着时间发散,导致轨道运动也随之发散。同时,Hamilton 形式对应的角动量残差始终保持为零,而 Lagrange 形式对应的角动量残差达到 10^7 量级。因此,在分析超大柔性空间结构的动力学特性时,使用 Hamilton 正则方程以及 SRK 法更为合理。考虑到姿态运动的大周期与柔性振动的小周期,数值仿真步长设置为 0.1 min 较为合适。接下来分别给出圆轨道和小偏心率轨道的重力梯度影响以及结构姿态耦合效应的仿真结果。

表 1 不同方程形式与数值算法组合的计算结果		
Tab. 1 Computational results of different combinations of formulations and algorithms		
方程与算法 组合形式	总角动量残差 最大值/ ($\text{N} \cdot \text{min} \cdot \text{km}$)	总能量残差 最大值/ ($\text{N} \cdot \text{km}$)
Hamilton & SRK	0	4.6×10^7
Hamilton & RK	0	1.4×10^{10} (发散)
Lagrange & SRK	6.5×10^7	9.2×10^7

3 圆轨道的动力学特性

3.1 重力梯度对轨道运动的影响

在不同初始姿态角 $\varphi(0)$ 情况下,选取结构尺寸 x_s 为 0.1 km , 1.0 km , 10.0 km 三种情况,用 RDwithGG 模型和 RDwithoutGG 模型响应之差表征重力梯度对轨道运动的影响。图 2(a) ~ (c) 和图 3(a) ~ (c) 分别给出了 $\varphi(0) = 0$ 时重力梯度对轨道半径和轨道角速度的影响曲线。两图中的 (a) ~ (c) 分别对应结构尺寸为 0.1 km , 1.0 km , 10.0 km 三种情况。从图 2(b) 和图 3(b) 可以观察到: $x_s = 1.0\text{ km}$ 时,轨道半径有向负向小幅波动,最大波动值为 $-2.5 \times 10^{-4}\text{ km}$;轨道角速度有向正向小幅波动,最大波动值为 $4.5 \times 10^{-9}\text{ rad/min}$,可见重力梯度对轨道运动影响的量级很小。同时,上述情况中轨道运动周期均为 88.4 min ,与不考虑重力梯度时的轨道运动周期相同,说明重力梯度对轨道周期没有影响。结构尺寸对重力梯度作用存在影响,从 RDwithGG 模型中的轨道运动方程可得,结构尺寸增大使得重力梯度增大。从图 2(a) ~ (c) 可知,随着结构尺寸的增大,重力梯度对轨道半径变化的影响按尺寸平方量级增大。从图 3 可以发现轨道角速度也有相同趋势。 $\varphi(0) = \pi/4\text{ rad}$ 时对应结构尺寸变化的影响与

$\varphi(0) = 0$ 时一致,也使重力梯度的影响按结构尺寸的平方量级增大。此外,注意到 RDwithGG 模型轨道运动方程中的重力项与姿态角有关,这说明重力梯度引起了轨道运动与姿态运动耦合,因此 $\varphi(0) = \pi/4\text{ rad}$ 中的结果出现不规则波动。对比 FDwithGG 模型和 FDwithoutGG 模型也得到相同结论。

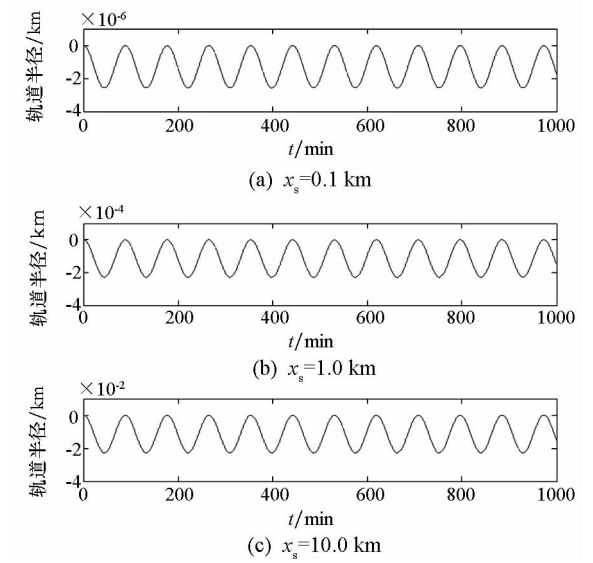


图 2 轨道半径变化($\varphi(0) = 0$)
Fig. 2 Change of orbital radius ($\varphi(0) = 0$)

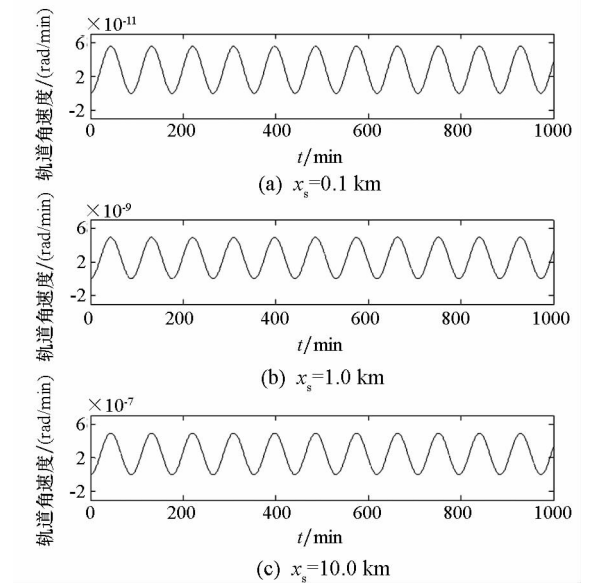


图 3 轨道角速度变化($\varphi(0) = 0$)
Fig. 3 Change of velocity of orbital angle ($\varphi(0) = 0$)

3.2 重力梯度对姿态运动的影响

在不同的初始姿态角 $\varphi(0)$ 情况下,选取结构尺寸 x_s 为 0.1 km , 1.0 km 两种情况,用姿态角和姿态角速度响应表征重力梯度对姿态运动的影响。在不考虑重力梯度时,从其对应的姿态运动方程可以看出, $\bar{m}x_s^2(\ddot{\varphi} + \ddot{\theta}) = p_\varphi = \text{const}$,即姿态

角动量为常数,这使得姿态运动只与轨道角速度以及初值有关。图4给出了对应的姿态角速度变

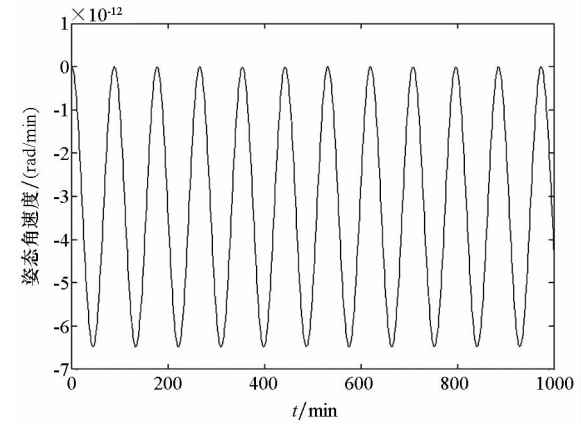
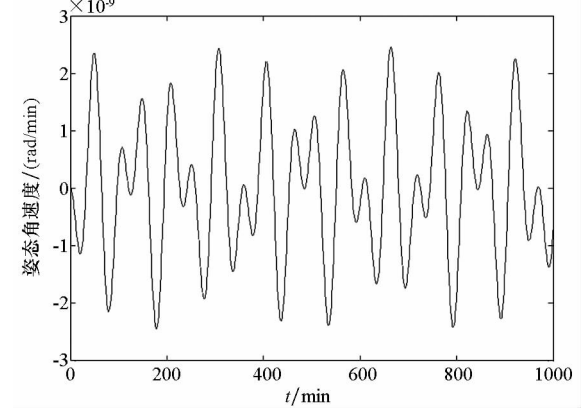


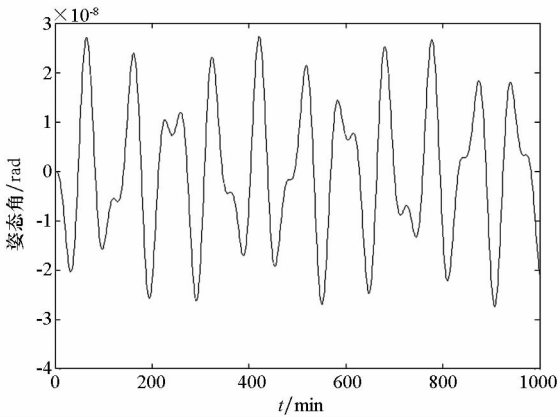
图4 不考虑重力梯度的姿态角速度变化
Fig.4 Change of velocity of attitude angle without effect of gravity gradient

化曲线,观察到姿态角速度一直为负,使得对应的姿态角始终反向运动,这说明不考虑重力梯度时姿态运动逐渐累积最后发散。而考虑重力梯度后,在对应姿态运动方程中可以看出姿态角动量受到重力梯度影响而不断变化,即姿态运动除了受轨道角速度影响,也受到重力梯度力矩的影响。图5(a)和图5(b)分别给出了 $\varphi(0)=0$ 时考虑重力梯度的姿态角速度和姿态角变化曲线,显示姿态角始终在平衡点附近振荡。可见轨道角速度的摄动使得姿态角从平衡点发散,而恢复力矩总是使姿态角回到平衡点,且影响量级相当。

图6(a)和图6(b)分别给出了 $\varphi(0)=0$ 时 x_s 为0.1 km和1.0 km时考虑重力梯度的姿态角响应。从图6中可以看出随着结构尺寸的增加,姿态角响应量级按尺寸平方的量级增加。在其他初始姿态角情况下,结构尺寸带来的影响略有不同。图7(a)和图7(b)分别给出了 $\varphi(0)=0.1$ rad时 x_s 为0.1 km和1.0 km时考虑重力梯度的姿态角

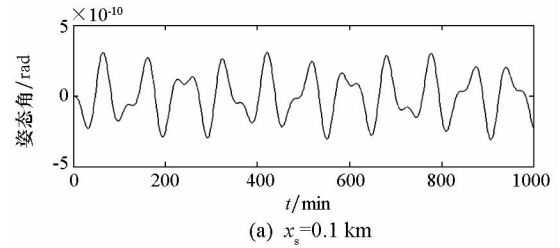


(a) 姿态角速度变化
(a) Change of velocity of attitude angle

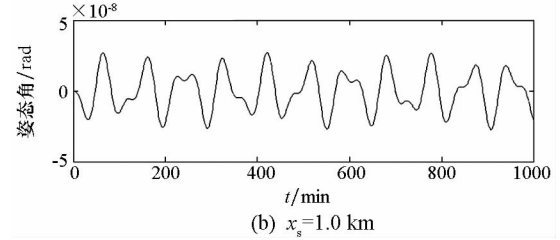


(b) 姿态角变化
(b) Change of attitude angle

图5 考虑重力梯度的姿态运动($\varphi(0)=0$)
Fig.5 Attitude motion with effect of gravity gradient ($\varphi(0)=0$)

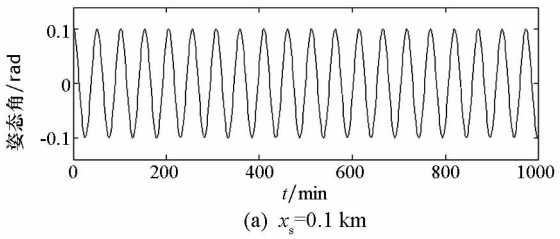


(a) $x_s=0.1$ km

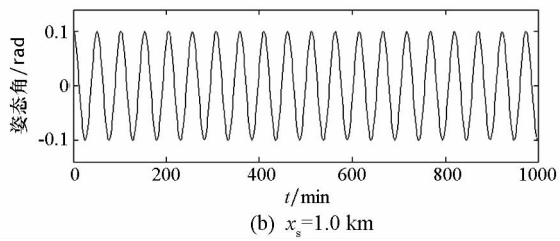


(b) $x_s=1.0$ km

图6 考虑重力梯度的姿态角响应($\varphi(0)=0$)
Fig.6 Attitude angle with effect of gravity gradient ($\varphi(0)=0$)



(a) $x_s=0.1$ km



(b) $x_s=1.0$ km

图7 考虑重力梯度的姿态角响应($\varphi(0)=0.1$ rad)
Fig.7 Attitude angle with effect of gravity gradient ($\varphi(0)=0.1$ rad)

响应。可以看出,随着尺寸的增加,姿态角响应几乎没有变化。注意到哑铃结构的转动惯量与尺寸的平方成正比,这就要求重力梯度产生的恢复力矩也与尺寸的平方成正比。对于 $\varphi(0) = \pi/4$ rad 的情况也有相同现象。

3.3 姿态运动和柔性振动的耦合效应

3.3.1 姿态运动对柔性振动的影响

在重力梯度的影响下,姿态运动与柔性振动之间出现耦合现象。图 8(a)~(c)分别给出了 $\varphi(0)$ 为 0, 0.1 rad, $\pi/4$ rad 时的结构变形量曲线,可以观察到曲线外部出现包络线,且包络线样式随初始姿态角变化而变化。在式(4)中的柔性振动方程中,右侧第一项为姿态运动项,右侧第二项为重力轴向分力项,可以看出重力轴向分力项与姿态角有关。

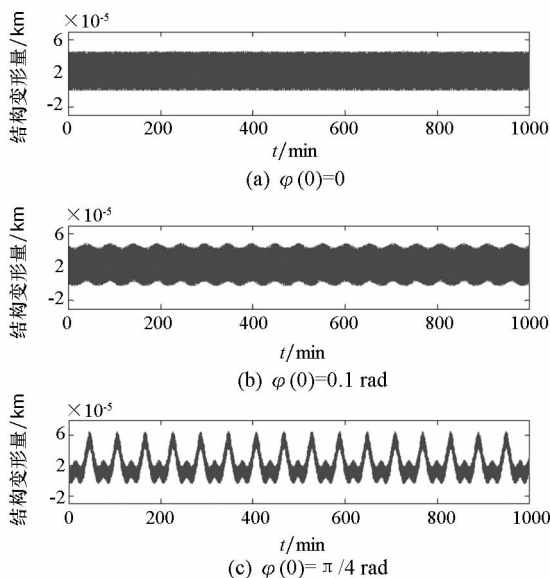


图 8 考虑重力梯度情况下的柔性振动响应

Fig. 8 Elastic vibration response with gravity gradient

图 9 给出了 $\varphi(0)$ 为 0.1 rad 时方程中部分项的数值变化曲线,其中虚线表示姿态运动项,实线表示重力轴向分力项,点划线表示两项之和。可以看出此时姿态运动项的变化幅度较大,重力轴向分力项的变化幅度很小,振动曲线的外部包络线样式与姿态运动项的变化样式一致。

图 10 给出了 $\varphi(0)$ 为 $\pi/4$ rad 时的结果。发现姿态运动项的变化幅度仍比重力轴向分力项的变化幅度大,但两者的数量级相同,在两者的共同作用下,出现了复杂的振动曲线外部包络线样式(如图 8(c)所示)。从图 9 和图 10 可以看出,重力梯度引起了姿态运动的变化,同时姿态运动影响了重力轴向分力,姿态运动与重力轴向分力共同影响了柔性振动曲线外部包络线样式,其中姿态运动起主要影响作用。

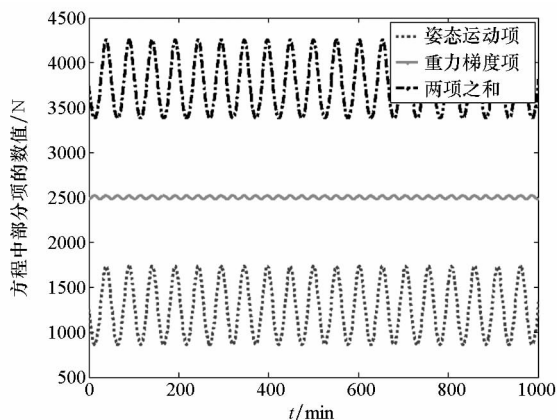


图 9 柔性振动方程中部分项变化曲线($\varphi(0) = 0$)

Fig. 9 Variation curve of several items in elastic vibration equation($\varphi(0) = 0$)

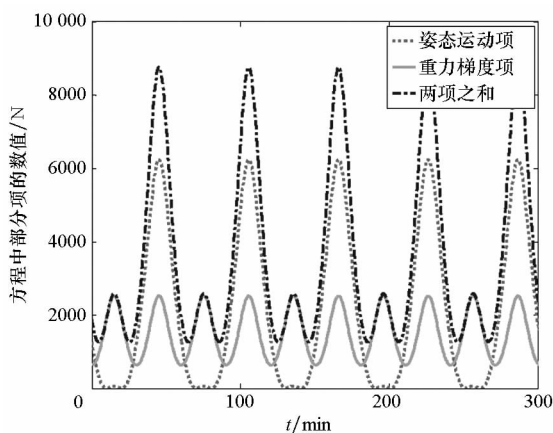


图 10 柔性振动方程中部分项随时间变化曲线($\varphi(0) = \pi/4$ rad)

Fig. 10 Variation curve of several items in elastic vibration equation($\varphi(0) = \pi/4$ rad)

3.3.2 柔性振动对姿态运动的影响

在重力梯度影响下,柔性振动对姿态运动也有影响,当尺寸较大时,这种影响尤为明显。图 11 给出了当 $k = 0.001k_s$, $\varphi(0) = 1.37$ rad 时考虑柔性与不考虑柔性的姿态角响应对比图,其中实线表示柔性体模型响应,虚线表示刚体模型响应。发现刚体模型的姿态运动周期保持不变,而柔性体模型的姿态运动周期持续变大。进一步增大结构的柔性振动,发现柔性振动对姿态运动的影响加强了。

图 12 分别给出了初始伸长量为 0 和 0.1 km (对应 $x(0)$ 分别为 1.0 km 和 1.1 km), $k = 0.01k_s$, $\varphi(0) = 1.26$ rad 时的结构变形量的姿态角响应曲线,其中实线表示初始伸长量为 0.1 km ($x(0) = 1.1$ km),虚线表示初始伸长量为 0 ($x(0) = 1.0$ km)。由图 12 可知,初始伸长量为 0.1 km 的姿态运动周期大于初始伸长量为 0 的姿态运动周期,这说明结构的柔性振动加剧后,柔性振动进一步增大了姿态运动的周期。

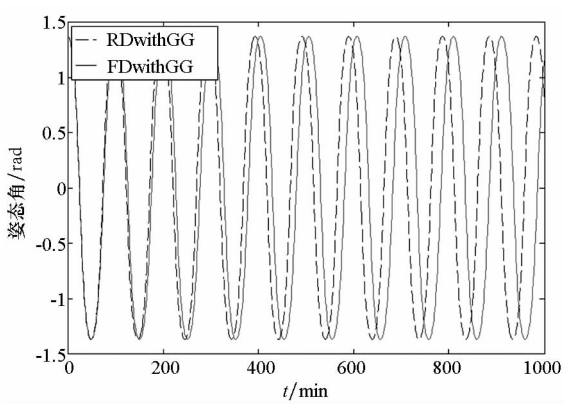


图 11 柔性体模型和刚体模型的姿态角
($\varphi(0) = 1.37 \text{ rad}$)

Fig. 11 Attitude angle of flexible and rigid model
($\varphi(0) = 1.37 \text{ rad}$)

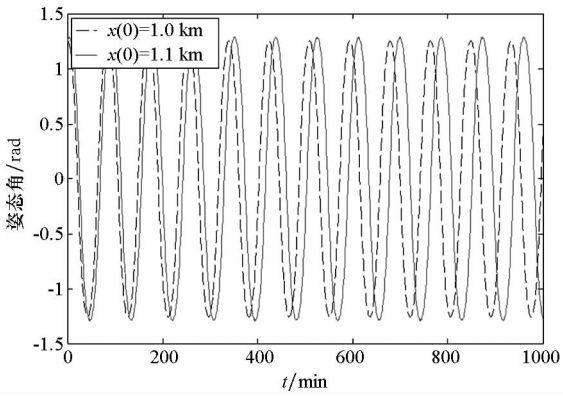


图 12 不同初始轴向变形量的姿态角

Fig. 12 Attitude angle with different initial
axial deformations

图 13 给出了在 $\varphi(0) = \pi/2 \text{ rad}$ 时柔性体模型和刚体模型的姿态运动,其中实线表示柔性体模型响应,虚线表示刚体模型响应。由图 13 可知,刚体模型始终保持在平衡位置附近,而柔性体模型则持续向负向翻滚,与 Malla 的结论^[5]一致,

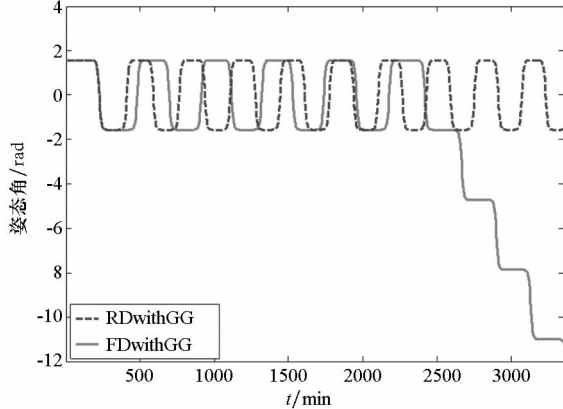


图 13 柔性体模型和刚体模型的姿态角
($\varphi(0) = \pi/2 \text{ rad}$)

Fig. 13 Attitude angle of flexible and rigid model
($\varphi(0) = \pi/2 \text{ rad}$)

即在重力梯度影响下,柔性振动使结构更容易发生翻滚现象。

4 小偏心率轨道的动力学特性

在小偏心率轨道下的结果基本与圆轨道下的结论一致,但是情况更为复杂。这主要是由于在小偏心率轨道下的轨道角速度变化幅度较大,其与重力梯度共同影响了姿态运动。图 14 给出了小偏心率轨道和圆轨道在 $\varphi(0) = 0.1 \text{ rad}$ 时姿态角响应对比,其中实线表示小偏心率轨道,虚线表示圆轨道,可以看到在小偏心率轨道下的姿态运动变得复杂。图 15(a) ~ (c) 分别给出了 $\varphi(0)$ 分别为 0, 0.1 rad 和 $\pi/4 \text{ rad}$ 时的结构变形量曲线。与圆轨道的结果(如图 8 所示)对比发现,柔性振动曲线的外部包络线样式变得复杂了。

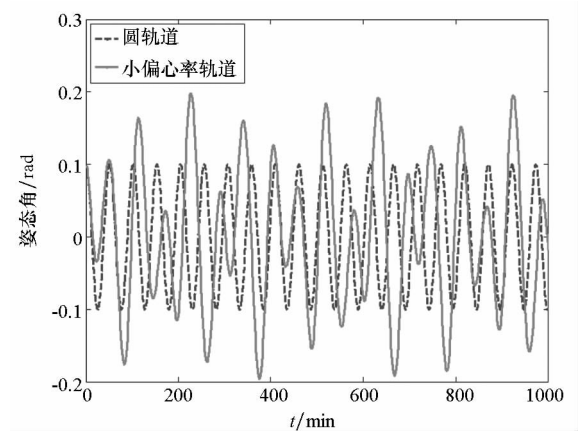


图 14 小偏心率轨道和圆轨道下的姿态角

Fig. 14 Attitude angle in circular orbit and
orbit with small eccentricity

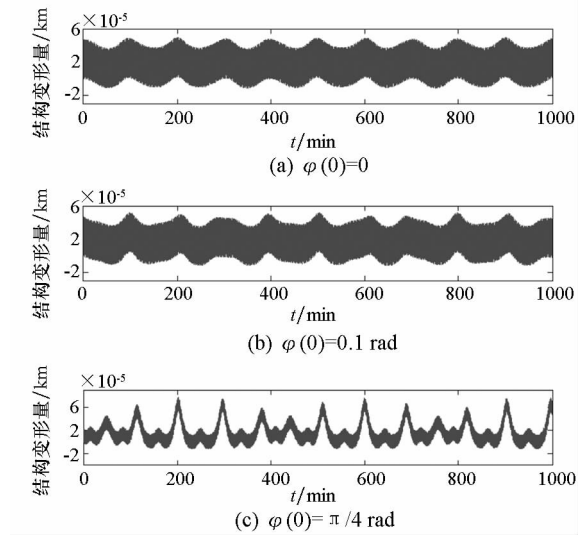


图 15 小偏心率轨道下的柔性振动响应

Fig. 15 Response of elastic vibration in
orbit with small eccentricity

此外,柔性振动对姿态运动的影响也有变化。

图 16 给出了在小偏心率轨道, $k = 0.01k_s$, $\varphi(0) = \pi/4$ rad 时, 柔性体模型和刚体模型的姿态角结果, 其中实线表示柔性体响应, 虚线表示刚体响应。不同于圆轨道在 $\varphi(0) = \pi/2$ rad 时发生翻滚, 小偏心率轨道在 $\varphi(0) = \pi/4$ rad 附近时发生翻滚。发生翻滚对应的初始姿态角称为临界初始姿态角。通过数值仿真, 图 17 给出了临界初始姿态角与轨道偏心率的变化曲线, 可以看到临界初始姿态角随轨道偏心率的增大呈近似于指数型的趋势减小。

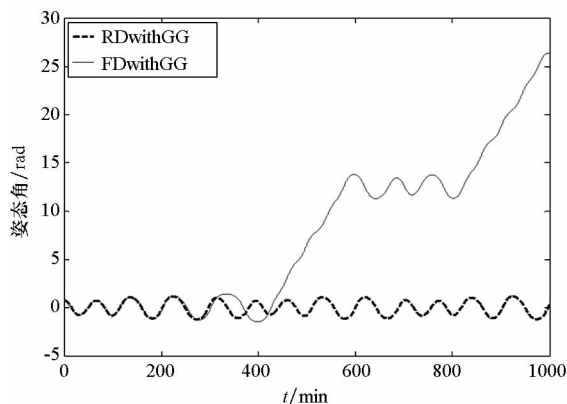


图 16 小偏心率轨道下柔性体和刚体模型的姿态角

Fig. 16 Attitude angle of flexible and rigid model in orbit with small eccentricity

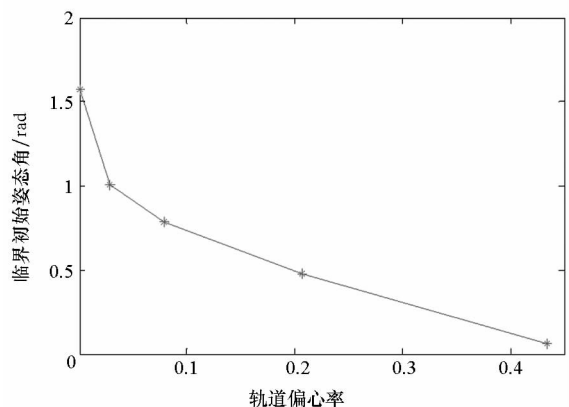


图 17 临界初始姿态角随轨道偏心率变化曲线

Fig. 17 Variation curve of critical initial attitude angle with orbit eccentricity

5 结论

主要结论如下:

1) 重力梯度对超大柔性空间结构的轨道运动影响较小, 但随着结构尺寸的增加, 重力梯度影响的量级按结构尺寸的平方量级增长;

2) 重力梯度力矩对超大柔性空间结构姿态动力学影响显著, 重力梯度力矩与结构尺寸的平方成正比;

3) 重力梯度使超大柔性空间结构振动与姿态运动产生耦合, 姿态运动会影响结构振动的模

式, 而结构振动则使得姿态运动周期增大, 并且结构振动使结构姿态运动更易于翻滚;

4) 小偏心率轨道下的结果与圆轨道下的结果基本一致, 但响应特性更加复杂。

目前的工作主要是采用数值仿真的手段揭示超大柔性空间结构在轨运行的特殊动力学现象, 接下来的工作将从理论上对重力梯度的影响方式、结构翻滚运动和稳定性等问题开展研究。

参考文献 (References)

- [1] Glaser P E. Power from the sun: its future [J]. Science, 1968, 162 (3856): 857 - 861.
- [2] Johnson W N, Bartolo R, Dorsey M, et al. Space-based solar power: possible defense applications and opportunities for NRL contributions: ADA513123 [R]. US: Naval Research Lab Washington DC Space Science DIV, 2009.
- [3] Belvin W K, Dorsey J T, Watson J J. Technology challenges and opportunities for very large in-space structural systems: LF99 - 9135 [R]. US: NASA Technical Reports Server, 2009.
- [4] 侯欣宾, 王立. 未来能源之路——太空发电站 [J]. 国际太空, 2014 (5): 4 - 7.
HOU Xinbin, WANG Li. Future energy path-solar power station [J]. Space International, 2014 (5): 4 - 7. (in Chinese)
- [5] Mankins J C. A fresh look at space solar power: new architectures, concepts and technologies [J]. Acta Astronautica, 1997, 41 (4/5/6/7/8/9/10): 347 - 359.
- [6] Sasaki S, Tanaka K, Higuchi K, et al. A new concept of solar power satellite: tethered - SPS [J]. Acta Astronautica, 2007, 60 (3): 153 - 165.
- [7] Mori M, Kagawa H, Saito Y. Summary of studies on space solar power systems of Japan aerospace exploration agency [J]. Acta Astronautica, 2006, 59 (1/2/3/4/5): 132 - 138.
- [8] 侯欣宾, 王立, 朱耀平, 等. 国际空间太阳能电站发展现状 [J]. 太阳能学报, 2009, 30 (10): 1443 - 1448.
HOU Xinbin, WANG Li, ZHU Yaoping, et al. International development situation of space solar power station [J]. Acta Energeticae Solaris Sinica, 2009, 30 (10): 1443 - 1448. (in Chinese)
- [9] Malla R B. Structural and orbital conditions on response of large space structures [J]. Journal of Aerospace Engineering, 1993, 6 (2): 115 - 132.
- [10] Ishimura K, Higuchi K. Coupling among pitch motion, axial vibration, and orbital motion of large space structures [J]. Journal of Aerospace Engineering, 2008, 21 (2): 61 - 71.
- [11] Sanyal A K, Shen J, McClamroch N H. Dynamics and control of an elastic dumbbell spacecraft in a central gravitational field [C] // Proceedings of 42nd IEEE Conference on Decision and Control, IEEE, 2003, 3: 2798 - 2803.
- [12] Sanyal A K, Shen J, McClamroch N H, et al. Stability and stabilization of relative equilibria of dumbbell bodies in central gravity [J]. Journal of Guidance, Control, and Dynamics, 2005, 28 (5): 833 - 842.
- [13] Ashley H. Observations on the dynamic behavior of large flexible bodies in orbit [J]. AIAA Journal, 1967, 5 (3): 460 - 469.
- [14] Sincarsin G B, Hughes P C. Gravitational orbit-attitude coupling for very large spacecraft [J]. Celestial Mechanics, 1983, 31 (2): 143 - 161.
- [15] Hairer E, Lubich C, Wanner G. Geometric numerical integration: structure-preserving algorithms for ordinary differential equations [M]. Springer Science & Business Media, 2006.

高超声速飞行器滑翔制导方法综述*

潘亮¹, 谢愈¹, 彭双春¹, 徐明亮², 袁天保³

(1. 国防科技大学 机电工程与自动化学院, 湖南 长沙 410073;
2. 北京跟踪与通信技术研究所, 北京 100094; 3. 火箭军装备研究院, 北京 100094)

摘要: 阐明高超声速飞行器滑翔制导的基本问题, 分析滑翔制导过程面临的复杂多约束、机动任务要求、参数扰动等研究难点; 分别就国内外标准轨迹制导方法和预测-校正制导方法相关研究现状展开综述, 指出了这两类方法中存在的问题。在此基础上, 提出高超声速飞行器滑翔制导研究中亟待解决的关键问题, 并指出未来滑翔制导方法的研究热点。

关键词: 高超声速飞行器; 滑翔制导; 标准轨迹制导; 预测-校正制导; 综述
中图分类号: V448 **文献标志码:** A **文章编号:** 1001-2486(2017)03-015-08

A survey of gliding guidance methods for hypersonic vehicles

PAN Liang¹, XIE Yu¹, PENG Shuangchun¹, XU Mingliang², YUAN Tianbao³

(1. College of Mechatronic Engineering and Automation, National University of Defense Technology, Changsha 410073, China;
2. Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China;
3. Equipment Academy of the Rocket Force, Beijing 100094, China)

Abstract: The basic problem of gliding guidance for hypersonic vehicles was proposed, and the difficulties of complicated multiple constraints, maneuver requirements, and parameter perturbation in the course of gliding guidance were analyzed. The corresponding research status at home and abroad was surveyed, and the problems were also pointed out. On this basis, the key problems required to be solved at present in the research of gliding guidance for hypersonic vehicles were presented, and the research hotspots in the methods of future gliding guidance were also pointed out.

Key words: hypersonic vehicle; gliding guidance; standard trajectory guidance; predictor-corrector guidance; survey

虽然近年来高超声速技术取得了较大进展, 但总体而言, 高超声速飞行技术相关理论研究还不够系统和深入, 离实际应用差距较大。制导系统是被誉为飞行器“大脑”的核心组成部分, 是保证飞行器平稳可靠飞行的关键系统。在高超声速飞行器滑翔机动飞行过程中, 飞行速度快、机动范围大、参数扰动强、气动/气动热等复杂飞行环境对飞行器的影响显著, 如何在滑翔机动制导中综合考虑这些因素, 是高超声速飞行器制导理论需要解决的关键科学问题。

本文研究的对象特指一种飞行速度超过 5 马赫、具有大升阻比气动外形的无动力高超声速飞行器。它通过助推火箭发射到一定高度或从空间轨道释放后, 利用气动升力在临近空间作长时间远距离高超声速滑翔飞行, 如美国的高超声速通用航天飞行器(Common Aero Vehicle, CAV)。

1 高超声速飞行器滑翔制导基本问题

制导是指导弹等制导武器在飞行过程中, 克服各种干扰因素, 使之按照选定的制导律或者预定的基准弹道, 导引武器飞向目标的过程^[1]。制导律描述的是制导武器接近目标的整个飞行过程所应遵循的运动规律^[2]。本文研究的高超声速飞行器飞行过程可描述为: 由助推火箭发射或者从天基平台释放, 初始高度处在大气层边缘, 初始速度略低于第一宇宙速度。再入大气层后进行远距离无动力滑翔飞行, 滑翔过程中要满足驻点热流、过载、动压、控制等多种约束条件限制。在接近目标上空时, 转入俯冲段, 进行快速转弯俯冲, 进而释放有效载荷对目标进行打击。美国高超声速 CAV^[3-4] 载荷释放过程和所携带的典型有效载荷分别如图 1

* 收稿日期: 2016-01-10
基金项目: 国家自然科学基金资助项目(11502289)
作者简介: 潘亮(1973—), 男, 新疆库尔勒人, 副研究员, 博士, E-mail: panliang.2000@gmail.com;
谢愈(通信作者), 男, 副研究员, 博士, E-mail: 15111155218@139.com

和图 2 所示。

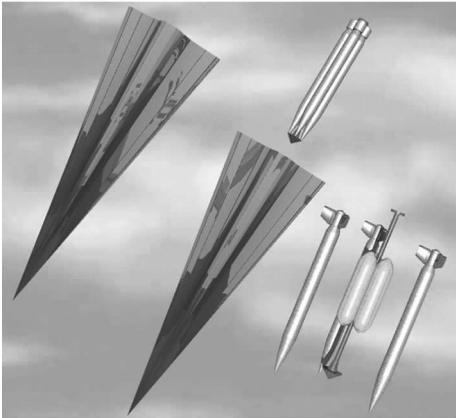


图 1 CAV 载荷释放过程
Fig. 1 Release of CAV payloads



图 2 CAV 有效载荷
Fig. 2 Payloads of CAV

滑翔飞行段的制导目的是综合考虑诸多复杂约束条件以及性能指标需求,控制飞行器在安全飞行走廊内完成远距离滑翔飞行,满足各种气动、气动热、过载等约束,实现飞行器高精度、强鲁棒性和强适应性制导的任务要求。由于飞行过程气动动态变化、气动热环境要求苛刻、热烧蚀不对称、飞行器模态变化剧烈,对飞行器制导控制提出了很高的要求。

滑翔飞行段制导问题可归纳为:从滑翔起点开始,在满足各种复杂过程约束的条件下,引导飞行器进行远距离无动力滑翔飞行,最终到达预定的终端点(或与俯冲段的交班点),并满足要求的终端或交班状态。滑翔段制导实际上与航天飞机、可重复使用运载器(Reusable Launch Vehicle, RLV)再入制导具有一定相似之处,但所受的约束条件更加苛刻,终端约束更强,对适应任务的灵活性要求更高。

2 高超声速飞行器滑翔机动制导问题分析

2.1 复杂多约束

高超声速飞行器长时间在临近空间高超声速飞行,既要满足设定的飞行任务要求,还要受到诸多复杂条件的约束,这些条件是进行制导设计时必须考虑的。根据约束条件的特点,可将其分为过程约束和端点约束两大类,如表 1 所示。

表 1 主要约束条件
Tab. 1 Major constraints

约束条件		主要目的
过程约束	热流密度约束	防止气动热烧蚀破坏
	动压约束	防止空气舵铰链力矩过大
	过载约束	防止飞行器结构性破坏
	平衡滑翔约束	保证飞行器平稳飞行
	控制量约束	满足飞行器控制能力要求
端点约束	起始条件	满足助推段与滑翔段交班条件
	终端约束	满足滑翔段和俯冲段交班条件

2.1.1 过程约束

热流密度约束:在研究飞行器制导问题时,通常以驻点热流密度峰值作为约束条件,因为驻点是飞行器加热较严重的区域。

动压约束:在飞行力学问题中,动压是最重要的特征量之一。所有气动力和力矩都与动压成比例,考虑到动压对控制系统的影响和侧向稳定性的要求,再入过程中的动压不能超过飞行器所能承受的极限值。

过载约束:为了结构安全,需要考虑过载约束。考虑到该研究对象为升力体机动滑翔飞行器,其轴向、法向都可能产生较大过载,影响飞行器结构的安全性,因此一般考虑对总过载进行限制。

平衡滑翔约束:平衡滑翔约束是一种考虑飞行器控制能力的“软约束”,即不是严格要求满足的约束条件,主要用于保证飞行器飞行过程平稳。

控制量约束:控制量通常取攻角和倾侧角,对控制量约束主要限制控制量的幅值和变化率。

2.1.2 端点约束

起始条件:再入起始点状态需要满足助推段与滑翔段交班状态要求,并决定了滑翔飞行器的飞行能力。再入起始点状态参数由助推段飞行状态决定。

终端约束:终端约束是为了保证滑翔段和俯冲攻击段交接班而对飞行器提出的状态要求。一般包括终端速度大小、终端高度、终端方位角、终端速度倾角等。根据交接班任务的不同,终端约束条件也不一样。

2.2 机动任务要求

高超声速飞行器在提高自身机动能力的同时,也增加了制导设计的难度。在进行制导方法研究时,需要重点考虑不同机动任务的要求。

例如:

1) 预设航路点:航路点是飞行器为满足多载荷释放或侦测等任务要求而需飞过的位置,一般要求飞行器从各航路点正上方飞过。

2) 目标变更:高超声速飞行器在实施飞行任务之前一般会预设目标点。但由于突发威胁、任务临时改变等原因,在飞行过程中可能要求飞行器在机动能力允许的情况下调整到新的目标点。

3) 应急处置:高超声速飞行器可能具备人在回路的控制能力。在飞行过程中,由于某些原因临时取消飞行任务,需要将飞行器引导到预先设定的应急处置区域。

4) 禁/避飞区:飞行器飞行过程中不允许或尽量避免经过某些区域,如敌方反导系统拦截区域、易受敌探测或电磁干扰的空域、地缘政治因素不允许通过的区域等。

2.3 参数扰动

在高超声速飞行器滑翔机动制导过程中,飞行高度和速度变化剧烈,气动力、气动热特性复杂,其制导控制系统不可避免地会受到各种扰动的影响,如大气密度、气动参数、飞行器质量、初始再入状态等均不同程度地存在扰动。由于传统的飞行走廊和参考飞行剖面设计过程中通常没有考虑参数扰动的影响,或者所考虑的扰动与实际情况存在较大差异,而飞行器在实际飞行过程中不可避免地会受到各种扰动因素的影响。虽然可通过增加反馈的形式将实际飞行剖面控制在参考剖面附近,但当设计的飞行剖面位于靠近走廊边界的位置时,仍有可能使得实际飞行约束超出最大允许值。此外,在这些干扰和摄动的作用下,使得制导系统参数存在扰动,制导信息往往不完全是平稳信号。在这种情况下,制导系统的稳定性和鲁棒性会降低,这将直接影响到制导效果,甚至导致高超声速飞行器滑翔机动制导过程出现失稳现象。

可见,高超声速飞行器滑翔制导问题是一个复杂多约束不确定性系统的鲁棒控制问题,需要重点考虑多约束影响、机动任务要求、参数扰动等情况。因此,研究先进滑翔制导理论和方法,保证飞行器平稳飞行,并充分发挥其机动能力优势,提高制导控制系统的鲁棒性,对于研发高超声速飞行器具有十分重要的理论和实践意义。

3 滑翔制导方法相关研究现状综述

3.1 国内外研究现状

3.1.1 标准轨迹制导

标准轨迹制导,即通过对预先设计好的参考

轨迹进行跟踪来实现制导。标准轨迹制导通常可分为标准轨迹设计和标准轨迹跟踪两个主要步骤。

标准轨迹制导在美国航天飞机返回再入制导中得到了成功应用,在飞行器再入制导方法中具有重要地位,它已成为了 X-33、X-37 等许多再入飞行器的基准方法。该方法主要包括离线参考阻力加速度剖面规划和在线阻力加速度剖面跟踪。参考阻力加速度剖面一般包括两个二次曲线段、一个拟平衡滑翔段、一个常阻力段和一个线性能量段。在规划过程中需将剖面限制在飞行走廊内,以满足航程要求和气动热、过载、动压、平衡滑翔等飞行约束条件;在线阻力加速度剖面跟踪,主要通过调整倾侧角跟踪参考阻力加速度剖面,并在倾侧翻转时对攻角进行微调,以尽量降低倾侧翻转过程带来的影响。当预测航程与期望航程存在偏差时,对参考阻力加速度剖面进行动态调整,以消除航程误差;航向通过基于侧向方位误差走廊的倾侧翻转进行控制^[5]。

为了适应新一代 RLV 再入制导的需要,产生了一系列再入制导方法的改进方案,主要体现在改进参考飞行剖面设计及更新方法、改进剖面跟踪控制方法等方面。

1) 改进剖面设计及更新方法

在经典的标准轨迹制导方案中,假设飞行器沿大圆弧飞行,并在二维再入飞行走廊内设计参考阻力加速度飞行剖面,没有考虑侧向机动需求,难以满足大范围侧向机动再入制导要求。为此,Mease 等^[6]在参考阻力加速度飞行剖面规划中同时考虑了纵向运动和侧向运动。该方法可以看作是航天飞机阻力剖面规划技术的自然扩展,分为轨迹长度规划子问题和轨迹曲率规划子问题。两个子问题需要反复迭代,最终确定出所需的三维轨迹。文献[7]在此基础上提出了一种演化的加速度再入制导方法(Evolved Acceleration Guidance Logic for Entry, EAGLE),并对其性能进行分析。EAGLE 最突出的特性在于能够规划三维轨迹,从而具备处理大侧向机动的再入制导问题的能力。文献[8]对 EAGLE 的性能做了进一步分析。文献[9]在以往 EAGLE 基础上,提出一种新的轨迹规划方法,能够获得近似的最大纵程和最大横程。

2) 改进剖面跟踪控制方法

20 世纪 90 年代初,Mease 等^[10-11]利用非线性几何中的反馈线性化理论,给出了基于阻力加速度剖面的轨迹跟踪方法。通过对跟踪控制律的理论分析表明,基于反馈线性化的跟踪制导更具

一般性。以此为基础,文献[12]进一步给出了滑模观测器,用来估计干扰,与反馈线性化跟踪制导配合使用。除了用于跟踪阻力加速度剖面外,作为经典制导方法的扩展,反馈线性化跟踪制导还用于跟踪其他形式的标准轨迹^[13-14]。文献[14]以乘员探索飞行器和军用航天飞机为背景,将二维制导扩展为完整的三维再入制导,并且将攻角和倾侧角均作为控制量,跟踪以能量为自变量的地面航迹。

除了反馈线性化跟踪制导外,另一种非线性跟踪制导律为基于预测控制理论的跟踪制导。Lu^[15]采用与经典的标准轨迹制导类似的制导方案,但参考阻力加速度剖面参数化为能量的分段线性函数并进行优化,提出了一种基于非线性预测控制的轨迹控制律。文献[16]进一步将基于预测控制的跟踪制导应用到 X-33 中,并给出了按照预测射程进行阻力剖面更新的方法。文献[17]在 Lu 所提方法基础上,选取不同的性能指标,最终获得了与反馈线性化跟踪制导不同的制导律,其优势在于能够适用于控制量饱和的情况。

前面所述方法大都属于基于阻力加速度剖面的跟踪制导,另一种思路是将轨迹跟踪问题处理为在参考轨迹状态空间考虑的调节问题。为解决该线性时变系统带来的难题, Lu^[18]基于滚动时域控制方法近似求解,给出了一种闭环稳定控制律。文献[19]同样将再入制导作为在状态空间考虑的相对参考轨迹的轨迹调节问题,并分别以能量和待飞航程为独立变量进行分析。Dukeman^[20]基于线性二次型调节器给出了一种标准轨迹跟踪制导方法。标准轨迹参数包括参考状态(待飞航程、高度和飞行路径角)和参考控制量(攻角和倾侧角)。纵向采用跟踪制导方式,侧向制导则仍通过与经典的标准轨迹制导类似的基于方位误差走廊的倾侧翻转实现。

3.1.2 预测-校正制导

标准轨迹制导方法虽然在工程实际中得到了应用,但是该方法必须预先规划参考轨迹,灵活性不足,且对初始再入条件敏感,无法完全满足新一代再入飞行器或高超声速飞行器制导要求。为此,人们一直在追求具有自主能力的预测-校正制导方法。预测-校正制导方法是以消除实际轨道的预报落点和预定落点位置之间的偏差为目的的制导方法。与标准轨迹制导方法不同,预测-校正制导不依赖于标准轨迹,而是在飞行过程中对终端状态不断进行预测,根据与期望终端状态

的偏差校正控制量。因此,该方法包括两个步骤:①基于实际飞行状态进行终端状态的快速预测;②根据预测结果进行弹道控制量的校正。按照预测方法的不同,预测-校正制导方法可分为解析预测-校正制导和数值预测-校正制导。

1) 解析预测-校正制导

解析预测-校正制导基本原理为:通过将轨迹调制到特定形式而获得轨迹的近似解析解,在每一制导周期中对飞行器终端状态进行解析预测,根据预测的终端状态偏差校正控制量。

从 20 世纪 80 年代开始,解析预测-校正方法就开始在大气捕获及火星探测任务的弹道规划与制导研究中受到重视并得到快速发展^[21-25]。针对火星探测中大气捕获和精确着陆问题, Bryant 等^[21]基于参考阻力-高度变率提出了一种解析的阻力控制算法;而 Masciarelli 等^[22]同样也是基于参考高度变率和阻力的计算,开发了一种解析预测-校正算法以满足火星返回轨道设计要求。Hanak 等^[23]则对这种解析预测-校正算法进行了改进,使其具备了更强的适应能力。与前述方法不同, Lafontaine 和 Levesque 等^[24-25]在假设了火星为不旋转行星的情况下,针对火星精确着陆问题提出了一种新的解析预测-校正方法,利用常飞行路径角或者飞行路径角与大气密度成比例关系的弹道剖面进行终端状态的解析预测。此外,还有部分学者研究基于平衡滑翔假设的弹道解析预测-校正方法。如 Tigges 等^[26]利用平衡滑翔条件实现了火星再入弹道的快速预测,提出了一种解析预测-校正的再入弹道快速生成与制导方法;而 Xu 等^[27]则以大升阻比高超声速滑翔飞行器为研究对象,利用平衡滑翔条件提出了一种自适应预测-校正的弹道快速生成方法。该方法首先根据控制终端速度及消除终端航向误差的需求实时确定倾侧角,将飞行器速度方向调整到射面内,进而基于拟平衡滑翔条件通过对终端状态的预测来调整攻角。

由于解析预测-校正制导方法采用解析公式进行在线轨迹预测计算,故计算量和存储量均很小,便于工程应用。

2) 数值预测-校正制导

数值预测-校正制导基本原理为:通过对整个飞行过程中的控制量进行参数化,使得控制量序列可由几个待定的控制参数进行描述,在飞行过程中利用运动方程的数值积分对终端状态进行预测,根据终端偏差来校正控制参数。

由于数值预测-校正制导需要实时轨迹数值

积分,故在线计算量很大。随着计算机技术的不断提高,人们逐渐开始重视数值预测-校正方法,研究重点包括如何降低数值预测的计算量以及控制量的校正方法等。为解决救生返回制导面临的不确定性问题,Powell^[28]为空间站的返回救生船设计了一种预测-校正制导律。此外,Powell还为火星探测器设计了一种典型的滚转角翻转数值预测-校正再入制导律^[29]。其中,控制量由滚转角幅值和翻转时间两个参数描述,弹道预测过程采用四阶Runge Kutta积分,而控制量校正则采用了二分法。Fuhry^[30]和Lu^[31]分别针对再入问题,设计了相应的预测-校正方法。Youssef^[32]针对RLV、X-33等再入飞行器,研究了数值预测-校正方法,控制参数可以包括倾侧角和攻角以及切换时间等多种组合,主要验证了取不同控制参数时的制导效果。国内学者针对控制量校正问题提出了采用单纯形法^[33]、模糊逻辑^[34]等方法。为降低数值预测-校正方法的计算量,Xu等^[35]基于神经网络建模方法实现了弹道的快速预测。雍恩米^[36]利用预设航路点将再入弹道分段,每次只预测飞行器到下一个航路点而不是目标点的终端状态,从而降低在线弹道预测的计算量。Zhang等^[37]则提出了每个制导周期只计算一次校正而不进行迭代的思路。

3.2 研究现状评述

从实际应用的角度,无论是标准轨迹制导方法还是预测-校正制导方法都存在一些问题。

1) 标准轨迹制导法着重借鉴航天飞机取得的成功经验,在飞行走廊内设计解析形式的阻力加速度-速度剖面(或其他类型剖面)作为参考轨迹,满足过程约束和航程要求;通过设计侧向方位误差走廊来实现倾侧翻转,实现对航向的控制。在这类标准剖面再入制导方法研究中,侧向机动制导仅作为纵向制导的“陪衬”,并没有充分发挥高超声速飞行器的横侧向高机动能力,更没有考虑机动任务以及航路点、禁/避飞区等约束下的实际需求,因而标准轨迹制导法难以满足高超声速飞行器复杂动态的机动任务需求。

2) 预测-校正制导在飞行过程中对终端状态不断进行预测,根据所预测的终端状态与期望终端状态的偏差校正控制量。但无论是解析预测-校正制导还是数值预测-校正制导都有其难以克服的缺陷:

解析预测-校正制导通过将轨迹调制到特定形式而获得轨迹的近似解析解,在每一制导周期中对飞行器终端状态进行解析预测,根据预测的

终端状态偏差校正控制量。由于在线计算需要采用解析公式以及动力学模型的复杂性,往往只能求取近似解析解,预测模型误差大,制导精度低,并且缺少对严格飞行过程约束的处理能力,难以适应高超声速飞行器在多约束条件下有效执行机动任务的制导要求。同时,参数扰动情况将会进一步加剧预测模型误差,大大提高制导控制的难度。

数值预测-校正制导通过对整个飞行过程中的控制量进行参数化,在飞行过程中利用数值积分对终端状态进行预测,根据终端偏差来校正控制参数,高超声速滑翔飞行器飞行距离远、机动范围大、在线计算量很大,目前由于弹载计算机水平的限制,大量的在线实时计算量使其还难以直接在实际中应用。

此外,无论是解析预测-校正制导还是数值预测-校正制导都着重于纵向制导方法的设计,而忽略了侧向机动能力的预测与侧向机动弹道的规划,因而限制了高超声速飞行器侧向机动能力的发挥,难以适应飞行器在复杂多变的战场环境下进行各种机动任务的实际需求。

4 滑翔制导亟待解决的关键问题

4.1 机动能力快速预测模型构建

高超声速飞行器因其具有快速反应能力、强突防能力、高机动作战及精确打击能力,具有巨大的军事价值和潜在的经济价值,为了充分利用其机动能力优势,扩展其战场空间,构建基于三维剖面的机动能力预测模型,显得尤为重要和迫切。为了解决直接由动力学积分求解机动能力模型计算量大、求解时间长、存储数据多等问题,研究高超声速飞行器快速预测模型,提出从当前状态点对应的目标覆盖区域问题的解析计算方法,将复杂的动力学积分计算问题转化为基于平面曲线理论的弹道快速生成方法,值得深入研究和探讨。

4.2 不同机动任务情况下制导方法的自适应设计

在不同的作战任务要求下,高超声速飞行器的机动任务是不同的,即使执行一次作战任务,飞行器机动任务亦可能会因为某些因素(如突发威胁、目标变更)而改变,离线条件下设计的制导方法难以完全确保飞行器滑翔机动过程的稳定性和适应性。在这种情况下,如何通过对机动任务及形式进行分析、分类和抽象,构建不同的机动任务模型,快速设计合理的机动策略,以保证高超声速飞行器针对不同机动任务的实时性和自适应性,

是解决飞行器圆满完成作战任务的关键理论问题,需要结合机动任务模型特点和滑翔机动制导理论方法展开研究。

4.3 参数扰动情况下制导方法的鲁棒性设计

参数扰动会影响制导效果,甚至造成飞行器在飞行过程中的结构性损坏,是高超声速飞行器滑翔机动制导需要解决的基本问题。针对这类问题,深入分析不同参数的扰动偏差模型,构建考虑参数扰动情况下的飞行走廊,进而在飞行走廊内进行制导方法的优化设计,将是解决参数扰动情况下高超声速滑翔飞行器制导鲁棒性设计的关键理论问题。

5 未来制导方法研究热点

5.1 三维剖面制导方法研究

能够进行大幅度横侧向机动飞行,是高超声速飞行器相对于弹道导弹等传统武器的最突出优势之一。这使得飞行任务对横侧向机动能力需求越来越高。基于这一点,学者们在原有再入制导方法的基础上,针对航路点和禁飞区等复杂飞行任务问题都做了深入研究,比如 Jorris 等^[38]研究的考虑禁飞区和航路点的三维轨迹快速再入制导方法,Chen 等^[39]研究的满足约束条件下临近空间高超声速滑翔飞行器三维轨迹快速生成,Xie 等^[40]研究的考虑航路点和禁飞区的三维再入轨迹生成等。这些方法大多集中在基于特定航路点以及禁飞区等约束下三维轨迹生成方法,并不具备普适意义。同时这些方法一般采用事先给定的固定攻角方案,主要通过调节倾侧角进行侧向机动,并保证平衡滑翔飞行,控制任务繁重,而倾侧角本身的调节能力有限,因此,采用这类制导方法仍旧难以充分发挥飞行器横侧向机动能力。

基于三维剖面的制导方法是再入制导方法中一个相对前沿且新颖的研究领域,目前仍处于探索研究阶段。基于三维剖面的制导方法原理是:综合考虑纵向航程与侧向机动任务需求,产生满足各种约束条件下的三维剖面,通过跟踪剖面得到需要的攻角和倾侧角,从而控制飞行轨迹。从可查阅到的国外文献来看,Mease^[6,11]在三维剖面制导方法上的研究较为深入。在设计三维剖面时,Mease 将纵向航程和横程误差同时考虑进去以产生各种满足约束条件下的三维剖面,然后利用忽略纵向运动后得到降阶的三自由度动力学模型求解出所有的状态变量和控制量攻角和倾侧角。但该方法仅采用简单负反馈进行跟踪制导,

并未基于三维剖面进行制导方法的深入研究。后来,Mease 又研究了一种三维剖面制导方法,即改进的加速度再入制导方法。该方法中攻角仍旧采用事先给定的攻角速度函数,仅依靠调节倾侧角来控制再入轨迹。因此,倾侧角需要同时兼顾纵向航程与侧向横程误差,控制任务复杂繁重。国内以 Mease 所研究的三维剖面生成和 EAGLE 制导方法为蓝本,开展了相关研究。比如闫晓东^[41]设计了三维轨迹生成方法,郭继峰^[42]研究了三维自主再入制导方法等。这些研究工作为后续研究提供了良好的理论借鉴。

目前,虽然基于三维剖面的制导方法研究还处于初步阶段,但已体现了其特有的技术优势。与传统制导方法相比,基于三维剖面的制导方法主要有如下突出特点:

- 1) 增加侧向机动任务需求,从而需要设计的剖面维数增加,设计剖面时所需要的再入走廊将会变得更加复杂,从而增加了优化设计的难度;

- 2) 攻角不再采用事先给定的速度函数攻角方案,直接由三维剖面跟踪获得,通过攻角和倾侧角同时控制飞行轨迹,提高了飞行器的控制效率;

- 3) 将侧向机动任务需求考虑到剖面设计中,可有效提高横向机动控制能力,改善横程误差;

- 4) 通过跟踪三维剖面同时确定需要的攻角与倾侧角,既不会有控制量富余而造成多余的机动,也不会因为控制量不足而引起飞行过程的抖动,可有效减少能量浪费,增大滑翔距离,扩展可覆盖区域。

5.2 考虑飞行器动态特性的闭环预测制导

在实际制导过程中,飞行器是一个高阶的、存在惯性和阻尼的复杂系统。在制导方法研究过程中,按照研究惯例,基于瞬时平衡的基本假设,忽略了飞行器的动态效应以及飞行器系统模型的不精确性,必然给制导带来误差。对于临近空间高超声速飞行器而言,在极大动压的作用下,控制系统是否具备足够的控制效率和如何设计强应力环境下的控制系统,都是值得深入探讨的问题。而考虑复杂的高阶飞行器运动模型,又将给飞行器制导设计带来困难。今后的工作需要基于高维弹体运动模型进行制导律的数值解或解析表达式的推导。

同时,由于优化模型与实际飞行环境的不一致,如大气密度、高阶项引力摄动和控制偏差的影响,飞行器若仍按优化得到的控制曲线采用开环制导,则可能引起较大的再入偏差。因此,一方面,应定量分析各种扰动因素对飞行参数的影响;

另一方面,应研究有效的闭环制导方法或补偿措施以提高制导精度。随着计算机数值处理能力的提高,精度更高、鲁棒性更强、自主性更好的数值预测制导方法必将更具有实际应用的潜力。就目前的硬件水平而言,为了降低预测制导周期内的计算量,可以深入研究模型简化方法,并进行制导算法设计。

5.3 轨迹规划、制导与控制一体化

一直以来,制导与控制一体化是导弹系统研究的热点问题,文献[43]验证了制导姿控一体化设计对拦截高机动目标的优越性。文献[44]阐述了空间轨迹与空间姿态控制的关系,即飞行器的位形空间就是任务空间,对其进行控制的基本要求是可以打击目标,同时需要兼顾飞行器的姿态控制。轨迹规划的目标是根据战场态势生成能够连接起始点至目标点的飞行器运动轨迹;制导的任务是基于测量信息实时生成飞行器的期望加速度,保证飞行器逐渐接近目标;而控制器则能实现制导律要求的期望加速度,完成控制任务。如果能够建立飞行器规划、制导与控制的一体化模型,以舵偏角为输入,以飞行轨迹为输出,设计统一的控制律,将推动飞行器规划、制导与控制领域的革新。

6 结论

本文围绕高超声速飞行器滑翔制导问题展开综述,探讨了高超声速飞行器滑翔制导面临的突出问题;论述国内外相关研究概况与存在的问题;并提出亟待解决的关键科学问题,指出了未来制导方法的研究热点。本文的研究工作,对于解决高超声速飞行器滑翔制导的关键理论问题具有参考意义。

参考文献 (References)

- [1] 羌缪. 导弹技术词典: 导弹系统[M]. 北京: 中国宇航出版社, 1991: 188-251.
QIANG Liu. Dictionary of missile technology: missile system[M]. Beijing: China Astronautic Publishing House, 1991: 188-251. (in Chinese)
- [2] 徐明友. 弹箭飞行动力学[M]. 北京: 国防工业出版社, 2003: 99-106.
XU Mingyou. Flight dynamics of missiles and rockets[M]. Beijing: National Defense Industry Press, 2003: 99-106. (in Chinese)
- [3] Richie G. The common aero vehicle: space delivery system of the future[C]//Proceedings of Space Technology Conference and Exposition, AIAA99-4435, 1999.
- [4] Terry H P. A common aero vehicle (CAV): model, description, and employment guide [R]. Air Force Research

- Laboratory, 2003.
- [5] Harpold J C, Graves C A, Jr. Shuttle entry guidance[J]. Journal of the Astronautical Sciences, 1979, 27 (3): 239-268.
- [6] Mease K D, Chen D T, Schonenberger H, et al. Reduced-order entry trajectory planning for acceleration guidance[J]. Journal of Guidance, Control, and Dynamics, 2002, 25(2): 257-266.
- [7] Leavitt J, Saraf A, Chen D, et al. Performance of evolved acceleration guidance logic for entry [C]//Proceedings of AIAA Guidance, Navigation, and Control Conference and Exhibit, AIAA-2002-4456, 2002.
- [8] Saraf A, Leavitt J, Chen D, et al. Design and evaluation of an acceleration guidance algorithm for entry[C]//Proceedings of AIAA Guidance, Navigation, and Control Conference and Exhibit, AIAA-2003-5737, 2003.
- [9] Leavitt J A, Mease K D. Feasible trajectory generation for atmospheric entry guidance [J]. Journal of Guidance, Control, and Dynamics, 2007, 30(2): 473-481.
- [10] Mease K D. Shuttle entry guidance revisited [C]. AIAA Guidance, Navigation and Control Conference, Hilton Head Island, SC: AIAA, 1992.
- [11] Mease K, Kremer J P. Shuttle entry guidance revisited using nonlinear geometric methods [J]. Journal of Guidance, Control, and Dynamics, 1994, 17(6): 1350-1356.
- [12] Talole S E, Benito J, Mease K D. Sliding mode observer for drag tracking in entry guidance [C]//Proceedings of AIAA Guidance, Navigation and Control Conference and Exhibit, AIAA-2007-6851, 2007.
- [13] Sanjay B A, Mease K D. Tracking law for a new entry guidance concept [C]//Proceedings of 22nd Atmospheric Flight Mechanics Conference, AIAA-97-3581, 1997.
- [14] Bharadwaj S, Rao A V, Mease K D. Entry trajectory tracking law via feedback linearization [J]. Journal of Guidance, Control, and Dynamics, 1998, 21(5): 726.
- [15] Lu P. Entry guidance and trajectory control for reusable launch vehicles[J]. Journal of Guidance Control & Dynamics, 2012, 20(1): 143-149.
- [16] Lu P, Hanson J M, Bhargava S. An alternative entry guidance scheme for the X-33 [C]//Proceedings of 23rd Atmospheric Flight Mechanics Conference, AIAA-98-4255, 1998.
- [17] Benito J, Mease K D. Nonlinear predictive controller for drag tracking in entry guidance [C]//Proceedings of AIAA/AAS Astrodynamics Specialist Conference and Exhibit, AIAA-2008-7350, 2008.
- [18] Lu P. Regulation about time-varying trajectories-precision entry guidance illustrated [C]//Proceedings of Guidance, Navigation, and Control Conference and Exhibit, AIAA-99-4070, 1999.
- [19] Lu P, Shen Z, Dukeman G, et al. Entry guidance by trajectory regulation [C]//Proceedings of AIAA Guidance, Navigation, and Control Conference and Exhibit, AIAA-2000-3958, 2000.
- [20] Dukeman G A. Profile-following entry guidance using linear quadratic regulator theory [C]//Proceedings of AIAA Guidance, Navigation, and Control Conference and Exhibit, AIAA-2002-4457, 2002.
- [21] Bryant L E, Tigges M A, Ives D G. Analytic drag control for precision landing and aerocapture [C]//Proceedings of 23rd

- Atmospheric Flight Mechanics Conference, A98 - 37438, 1998.
- [22] Masciarelli J, Rousseau S, Fraysse H, et al. An analytic aerocapture guidance algorithm for the Mars sample return orbiter[C]//Proceedings of Atmospheric Flight Mechanics Conference, AIAA - 2000 - 4116, 2000.
- [23] Hanak C, Crain T, Masciarelli J. Revised algorithm for analytic predictor-corrector aerocapture guidance-exit phase [C]//Proceedings of AIAA Guidance, Navigation, and Control Conference and Exhibit Austin, AIAA - 2003 - 5746, 2003.
- [24] de Lafontaine J, Levesque J F, Kron A. Robust guidance and control algorithms using constant flight path angle for precision landing on Mars [C]// Proceedings of AIAA Guidance, Navigation, and Control Conference and Exhibit, AIAA 2006 - 6075, 2006.
- [25] Levesque J F, Lafontaine J D. Optimal guidance using density-proportional flight path angle profile for precision landing on Mars [C]//Proceedings of AIAA Guidance, Navigation, and Control Conference and Exhibit, AIAA 2006 - 6076, 2006.
- [26] Tigges M, Ling L. A predictive guidance algorithm for mars entry[C]//Proceedings of 27th Aerospace Sciences Meeting, AIAA - 89 - 0632, 1989.
- [27] Xu M L, Liu L H, Tang G J. Quasi-equilibrium glide auto-adaptive entry guidance based on ideology of predictor-corrector[C]//Proceedings of 5th International Conference on Recent Advances in Space Technologies, 2011.
- [28] Powell R W. Six-degree-of-freedom guidance and control entry analysis of the HL - 20 [J]. Journal of Spacecraft and Rockets, 1993, 30(5): 537 - 542.
- [29] Powell R W. Numerical roll reversal predictor-corrector aerocapture and precision landing guidance algorithm for the Mars surveyor program 2001 missions; AIAA - 98 - 4574[R]. NASA Langley Technical Report Server, 1998.
- [30] Fuhry D P. Adaptive atmospheric reentry guidance for the Kistler K - 1 orbital vehicle[C]//Proceedings of Guidance, Navigation, and Control Conference and Exhibit, AIAA - 99 - 4211, 1999.
- [31] Lu P. Predictor-corrector entry guidance for low lifting vehicles [C]//Proceedings of AIAA Guidance, Navigation and Control Conference and Exhibit, Hilton Head, South Carolina, 2008.
- [32] Youssef H, Chowdhry R S, Lee H, et al. Predictor-corrector entry guidance for reusable launch vehicles[C]//Proceedings of Guidance, Navigation, and Control Conference and Exhibit, AIAA - 2001 - 4043, 2001.
- [33] 呼卫军, 杨业, 周军. 基于外部信息源的临近空间飞行器中制导研究[J]. 航天控制, 2010, 28(2): 23 - 28.
HU Weijun, YANG Ye, ZHOU Jun. Research of midcourse guidance based on external information source for near space vehicle[J]. Aerospace Control, 2010, 28(2): 23 - 28. (in Chinese)
- [34] 王俊波, 曲鑫, 任章. 基于模糊逻辑的预测再入制导方法[J]. 北京航空航天大学学报, 2011, 37(1): 63 - 66.
WANG Junbo, QU Xin, REN Zhang. Predictive guidance method for the reentry vehicles based on fuzzy logic [J]. Journal of Beijing University of Aeronautics and Astronautics, 2011, 37(1): 63 - 66. (in Chinese)
- [35] Xu M L, Liu L H, Yang Y, et al. Neural network based predictor-corrector entry guidance for high lifting vehicles[C]// Proceedings of 62nd International Astronautical Congress, 2011.
- [36] 雍恩米. 高超声速滑翔式再入飞行器轨迹优化与制导方法研究[D]. 长沙:国防科技大学, 2008.
YONG Enmi. Study on trajectory optimization and guidance approach for hypersonic glide-reentry vehicle[D]. Changsha: National University of Defense Technology, 2008. (in Chinese)
- [37] Zhang Z, Hu J. Prediction-based guidance algorithm for high-lift reentry vehicles [J]. Science China Information Sciences, 2011, 54(3): 498 - 510.
- [38] Jorris T R, Cobb R G. Three-dimensional trajectory optimization satisfying waypoint and no-fly zone constraints [J]. Journal of Guidance, Control, and Dynamics, 2009, 32(2): 551 - 572.
- [39] Dong C, Chao T, Wang S Y, et al. Rapid three-dimensional constrained trajectory generation for near space hypersonic vehicles[C]//Proceedings of 18th AIAA/3AF International Space Planes and Hypersonic Systems and Technologies Conference, AIAA 2012 - 5896, 2012.
- [40] Xie Y, Liu L H, Liu J, et al. Rapid generation of entry trajectories with waypoint and no-fly zone constraints [J]. ACTA Astronautica, 2012, 77: 167 - 181.
- [41] 闫晓东, 王智. 高超声速无动力滑翔三维轨迹规划方法[J]. 北京理工大学学报, 2013, 33(7): 669 - 674.
YAN Xiaodong, WANG Zhi. Three-dimensional trajectory planning method for hypersonic glide vehicles [J]. Transactions of Beijing Institute of Technology, 2013, 33(7): 669 - 674. (in Chinese)
- [42] 郭继峰, 傅瑜, 崔乃刚. 三维自主再入制导方法[J]. 控制与决策, 2013, 28(5): 688 - 694.
GUO Jifeng, FU Yu, CUI Naigang. Three dimensional autonomous entry guidance method [J]. Control and Decision, 2013, 28(5): 688 - 694. (in Chinese)
- [43] Zhurbal A, Idan M. Effect of estimation on the performance of an integrated missile guidance and control system [C]//Proceedings of Guidance, Navigation and Control Conference and Exhibit, AIAA 2008 - 7458, 2008.
- [44] 韩大鹏. 基于四元数代数和李群框架的任务空间控制方法研究[D]. 长沙:国防科技大学, 2008.
HAN Dapeng. Research on task-space control based on quaternion algebra and a lie-group framework[D]. Changsha: National University of Defense Technology, 2008. (in Chinese)

高超声速再入轨迹跟踪控制的微分变换方法*

刘莉^{1,2}, 杨乐平¹, 蔡伟伟³

(1. 国防科技大学 航天科学与工程学院, 湖南 长沙 410073; 2. 空间物理重点实验室, 北京 100076;
3. 国防科技大学 指挥军官基础教育学院, 湖南 长沙 410073)

摘要:针对多约束条件下高超声速飞行器再入制导问题,提出一种基于微分变换法求解最优反馈控制的全状态标准轨迹跟踪制导律。利用滚动时域控制方法设计易于在线执行的闭环跟踪制导策略,在每个制导周期内将标准轨迹跟踪问题转化为线性时变系统状态调节器问题,并通过最优控制理论进一步转化为两点边值问题,采用微分变换法进行求解获得最优反馈控制律。数值仿真表明微分变换法的引入有效解决了传统两点边值问题求解的数值不稳定性与耗时问题,所设计的闭环制导律对状态偏差与模型不确定性具有较强的鲁棒性,可为工程设计提供有益参考。

关键词:微分变换;滚动时域控制;标准轨迹;再入制导

中图分类号:V488.2 文献标志码:A 文章编号:1001-2486(2017)03-023-07

Differential transformation-based trajectory tracking guidance scheme for hypersonic reentry vehicle

LIU Li^{1,2}, YANG Leping¹, CAI Weiwei³

(1. College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China;
2. Science and Technology on Space Physics Laboratory, Beijing 100076, China;
3. College of Basic Education, National University of Defense Technology, Changsha 410073, China)

Abstract: Concentrating on the hypersonic reentry guidance under multiple constraints, a full-state nominal trajectory tracking guidance scheme was proposed by applying the differential transformation approach to the optimal feedback control. In the period of the online closed-loop guidance scheme based on the receding-horizon control, the nominal trajectory tracking problem was transformed into a state regulator problem of the associated linear time-varying system, and then into a two-point boundary value problem by utilizing the optimal control theory. The differential transformation approach was suggested for the optimal feedback control, avoiding the time-consuming and numerical instabilities of conventional methods. Numerical simulation results validate that the proposed guidance scheme is robust to state dispersions and model uncertainties, providing a reference for engineering design.

Key words: differential transformation; receding-horizon control; nominal trajectory; reentry guidance

飞行器以高超声速再入地球大气层飞行时,面临严峻的气动力热环境,给飞行器结构、材料等带来巨大挑战。对高超声速飞行器而言,飞行制导是其安全飞行、成功遂行任务的有效支撑和重要保证,然而系统面临的强非线性动力学特性、复杂路径约束与控制约束等显著增大了再入制导律设计的难度,有必要开展深入研究。

按再入制导策略不同,常见的高超声速飞行制导律大致可分为预测-校正制导^[1-2]和标准轨迹制导^[3-4]两类。预测-校正制导通常包括两个步骤:一是在飞行过程中不断由飞行器当前状态积分预测终端状态,二是依据相对于期望终端状

态的偏差对制导指令进行调整。预测-校正制导按终端状态预测方法不同,可进一步分为解析预测-校正和数值预测-校正。前者预测速度快,但精度有限,且缺少对严格飞行约束的处理能力;后者精度较高,但数值积分计算量较大,导致预测速度较慢。

标准轨迹制导是在预先设计满足各类约束和任务要求的标准轨迹基础上,依据当前实际飞行轨迹相对于标准轨迹的偏差设计反馈控制律,确保飞行器沿标准轨迹飞行。为提高轨迹制导的鲁棒性和实时性,有关研究主要沿两方面展开:一是从轨迹规划方法着手,提高标准轨迹在线生成的

* 收稿日期:2016-08-31
基金项目:航空科学基金资助项目(2016ZC88007);中国运载火箭技术研究院高校联合创新基金资助项目(CALT201603)
作者简介:刘莉(1973—),女,吉林松原人,高级工程师,博士研究生,E-mail:437359021@qq.com

快速性;二是从轨迹跟踪算法着手,研究能在线实时解算且具有鲁棒性的跟踪算法。早期标准轨迹再入制导主要基于阻力加速度剖面的跟踪,并成功应用于航天飞机再入任务。虽然这种方式能够较好地控制航程以及终端能量,但对其余状态变量的控制能力有所欠缺。为此人们提出了基于状态空间的标准轨迹制导方法,将轨迹跟踪问题处理为状态调节问题来进行研究。针对线性时变系统的状态调节问题,近年来一种基于滚动时域控制方法的闭环制导策略被深入研究,并应用于再入制导^[3-4]、小推力轨道转移^[5]等领域,取得了较好的效果。该方法利用极大值原理将有限时域内的最优反馈控制问题转换为两点边值问题求解,但 Riccati 矩阵微分方程的传统求解方法存在耗时长、数值不稳定等不足。文献[6]基于 Legendre 伪谱法将两点边值问题推导出的线性时变方程转换为一系列离散线性代数方程求解。然而,该方法需进行大量高维矩阵运算,限制了其求解效率。近年来,微分变换法因其显著的求解效率与近似精度广泛应用于数值求解微积分方程^[7]。微分变换法实质是求微积分方程的泰勒级数解,但其在最优控制问题求解方面的应用并不多见。

1 问题描述

考虑地球为非旋转圆球,则半速度坐标系下的再入飞行器运动方程^[8]为:

$$\begin{cases} \dot{r} = V \sin \gamma \\ \dot{\theta} = V \cos \gamma \sin \psi / (r \cos \varphi) \\ \dot{\varphi} = V \cos \gamma \cos \psi / r \\ \dot{V} = -D/m - g \sin \gamma \\ \dot{\gamma} = \frac{1}{V} \left[\frac{L \cos \sigma}{m} + \left(\frac{V^2}{r} - g \right) \cos \gamma \right] \\ \dot{\psi} = \frac{1}{V} \left[\frac{L \sin \sigma}{m \cos \gamma} + \frac{V^2}{r} \cos \gamma \sin \psi \tan \varphi \right] \end{cases} \quad (1)$$

式中: r 、 θ 、 φ 和 V 分别为地心距、经度、纬度和速度;航迹倾角 γ 是速度矢量与当地水平面的夹角,向上为正;速度方位角 ψ 为速度向量在当地水平面投影与正北方向的夹角,顺时针旋转为正; σ 为倾侧角; L 、 D 分别表示升力和阻力,其表达式为

$$\begin{cases} L = \rho V^2 S_{\text{ref}} C_L / 2 \\ D = \rho V^2 S_{\text{ref}} C_D / 2 \end{cases} \quad (2)$$

式中: S_{ref} 为飞行器气动参考面积; ρ 为大气密度,

$$\rho = \rho_0 e^{-(r-R_0)/H_s} \quad (3)$$

其中 $H_s = 7110 \text{ m}$, $R_0 = 6378 \text{ km}$, $\rho_0 = 1.225 \text{ kg/m}^3$ 为海平面处大气密度。

再入飞行过程中,马赫数属于高超声速范围,气动系数近似满足阻力极线关系^[9]:

$$C_D = C_{D0} + K C_L^2 \quad (4)$$

式中零升阻力系数 C_{D0} 和诱导阻力因子 K 在高超声速下趋于常数。

定义飞行器泛化升力系数:

$$\lambda = C_L / C_L^* \quad (5)$$

式中, $C_L^* = \sqrt{C_{D0}/K}$ 为最大升阻比 E^* 对应的升力系数^[9],则飞行器的气动系数可表示为:

$$\begin{cases} C_L = \lambda C_L^* \\ C_D = \frac{C_L^*}{2E^*} (1 + \lambda^2) \end{cases} \quad (6)$$

对于指定飞行器,其气动特性参数 E^* 与 C_L^* 的取值均已知,故可将泛化升力系数 λ 作为弹道设计参数。

考虑飞行器的热防护、结构和控制性能,飞行过程中要求满足驻点热流密度 Q 、过载 n 、动压约束 q :

$$\begin{cases} Q = K_Q \dot{\rho}^{0.5} V^{3.15} \leq Q_{\max} \\ n = \sqrt{L^2 + D^2} / mg \leq n_{\max} \\ q = 0.5 \rho V^2 \leq q_{\max} \end{cases} \quad (7)$$

式中: K_Q 为常数,其取值与飞行器密切相关; Q_{\max} 、 n_{\max} 、 q_{\max} 分别为飞行器允许的最大驻点热流密度、最大过载和最大动压值。

2 闭环跟踪制导律设计

2.1 滚动时域控制策略

滚动时域控制策略如图1所示,其中 t_{EH} 为制导指令更新周期, $t_k (k=1, 2, \dots, n)$ 为制导指令切换时刻, t_{PH} 为滚动时域长度, t_p 为制导算法在线计算时间, u 为制导指令。滚动时域控制的基本思想是在有限时域 $[t_k, t_k + t_{\text{PH}}] (k=1, 2, \dots, n)$ 内,将动力学方程沿标准轨迹线性化,并以当前状态偏差为初始状态,构建该有限时域内的最优控制问题,求解获得反馈控制 $\mathbf{u}_{\text{opt}}(\tau) (t_k \leq \tau \leq t_k + t_{\text{PH}})$ 。值得注意的是,滚动时域控制仅选用 t_k 时刻的反馈控制 $\mathbf{u}_{\text{opt}}(t_k)$ 作为当前制导周期内的校正指令,所获得的其余反馈控制 $\mathbf{u}_{\text{opt}}(\tau) (t_k < \tau \leq t_k + t_{\text{PH}})$ 则全部舍弃。若反馈控制指令未能在规定时间内生成,则继续使用上一周期生成的指令;重复上述过程直至任务结束。滚动时域控制策略的可操作性强,且其闭环稳定性在控制理论上已经得到证明^[3]。

$$\Xi_x(t) = \sum_{i=0}^{\infty} X_{t_e}(i) \left(\frac{t-t_e}{H} \right)^i \tag{15}$$

将式(14)代入式(15)中,并取前 $N+1$ 项截断:

$$\begin{aligned} \Xi_x(t) &= \sum_{i=0}^N \left(\frac{\partial^i x}{\partial t^i} \right)_{t_e} \frac{(t-t_e)^i}{i!} + \left(\frac{\partial^{N+1} x}{\partial t^{N+1}} \right)_{t=\xi} \frac{(t-t_e)^{N+1}}{(N+1)!} \\ &\triangleq \sum_{i=0}^N \left(\frac{\partial^i x}{\partial t^i} \right)_{t_e} \frac{(t-t_e)^i}{i!} + R_N \end{aligned} \tag{16}$$

式中: ξ 为区间 $[t, t_e]$ 上的任意数; R_N 为泰勒展开定理的截断项,对充分大的 N 可忽略该项。对于光滑函数 $x(t)$,逆微分变换 $\Xi_x(t)$ 在任意点 $t \in (a, b)$ 处均能够收敛到原函数。

微分变换法求解一般微分方程 $\dot{x}(t)=f(x(t))$ 的过程一般包括三步:

1)在展开点 $t_e \in (a, b)$ 处利用微分变换将微分方程转换为一系列关于 $X_{t_e}(i)$ 的递推代数方程;

2)根据微分方程的边值确定 $X_{t_e}(0)$,并由递推方程计算出 $x(t)$ 的第 i 阶导数相应的微分变换值 $X_{t_e}(i) (i=1, \cdots, N)$;

3)将微分变换值 $X_{t_e}(i) (i=0, 1, \cdots, N)$ 代入式中,并忽略截断项,从而得到 N 阶泰勒展开形式的函数 $x(t)$ 近似解。

值得注意的是,微分方程与递推代数方程组间的转换依赖于微分变换的一些固有基本性质。表 1 给出了求解两点边值问题所需利用的性质,其中 $c \in \mathbb{R}$ 为常数。

表 1 微分变换的运算法则

Tab.1 Basic operations for differential transformation	
原函数	变换函数($i=0, 1, \cdots, N$)
$x(t) \pm y(t)$	$X_{t_e}(i) \pm Y_{t_e}(i)$
$c \cdot x(t)$	$c \cdot X_{t_e}(i)$
$dx(t)/dt$	$(i+1)X_{t_e}(i+1)/H$

3.2 求解两点边值问题

对于第 2.2 节给出的两点边值问题,由于其属于线性时变系统,因此在任意时刻 $t \in [t_{\text{now}}, t_{\text{end}}]$ 处其解可写为:

$$\begin{bmatrix} \Delta \mathbf{x}(t) \\ \Delta \boldsymbol{\lambda}(t) \end{bmatrix} = \begin{bmatrix} \mathbf{F}(t, t_{\text{end}}) & \mathbf{G}(t, t_{\text{end}}) \\ \mathbf{L}(t, t_{\text{end}}) & \mathbf{M}(t, t_{\text{end}}) \end{bmatrix} \begin{bmatrix} \Delta \mathbf{x}(t_{\text{end}}) \\ \Delta \boldsymbol{\lambda}(t_{\text{end}}) \end{bmatrix} \tag{17}$$

式中状态转移分块矩阵 $\mathbf{F}, \mathbf{G}, \mathbf{L}, \mathbf{M}$ 均为 6×6 维,且满足:

$$\begin{cases} \mathbf{F}(t_{\text{end}}, t_{\text{end}}) = \mathbf{I} \\ \mathbf{G}(t_{\text{end}}, t_{\text{end}}) = \mathbf{0} \\ \mathbf{L}(t_{\text{end}}, t_{\text{end}}) = \mathbf{0} \\ \mathbf{M}(t_{\text{end}}, t_{\text{end}}) = \mathbf{I} \end{cases} \tag{18}$$

将横截条件式(12)代入式(17)得:

$$\begin{cases} \Delta \mathbf{x}(t) = [\mathbf{F}(t, t_{\text{end}}) + \mathbf{G}(t, t_{\text{end}}) \mathbf{P}] \Delta \mathbf{x}(t_{\text{end}}) \\ \Delta \boldsymbol{\lambda}(t) = [\mathbf{L}(t, t_{\text{end}}) + \mathbf{M}(t, t_{\text{end}}) \mathbf{P}] \Delta \mathbf{x}(t_{\text{end}}) \end{cases} \tag{19}$$

定义矩阵 $\mathbf{V}(t)$ 和 $\mathbf{W}(t)$ 分别为:

$$\begin{cases} \mathbf{V}(t) = \mathbf{F}(t, t_{\text{end}}) + \mathbf{G}(t, t_{\text{end}}) \mathbf{P} \\ \mathbf{W}(t) = \mathbf{L}(t, t_{\text{end}}) + \mathbf{M}(t, t_{\text{end}}) \mathbf{P} \end{cases} \tag{20}$$

则协态矢量 $\Delta \boldsymbol{\lambda}(\cdot)$ 可表示为:

$$\Delta \boldsymbol{\lambda}(t) = \mathbf{W}(t) \mathbf{V}^{-1}(t) \Delta \mathbf{x}(t) \tag{21}$$

将式(19)~(21)代入式(10)中得:

$$\begin{cases} \dot{\mathbf{V}}(t) = \mathbf{A} \mathbf{V}(t) - \mathbf{B} \mathbf{R}^{-1} \mathbf{B}^T \mathbf{W}(t) \\ \dot{\mathbf{W}}(t) = -\mathbf{Q} \mathbf{V}(t) - \mathbf{A}^T \mathbf{W}(t) \end{cases} \tag{22}$$

相应边界条件为:

$$\begin{cases} \mathbf{V}(t_{\text{end}}) = \mathbf{I} \\ \mathbf{W}(t_{\text{end}}) = \mathbf{P} \end{cases} \tag{23}$$

运用微分变换方法,式(22)在终端时刻 t_{end} 处可变换为下列递归代数方程组:

$$\begin{cases} (k+1) \hat{\mathbf{V}}_{t_{\text{end}}}(k+1) = \mathbf{A} \hat{\mathbf{V}}_{t_{\text{end}}}(k) - \mathbf{B} \mathbf{R}^{-1} \mathbf{B}^T \hat{\mathbf{W}}_{t_{\text{end}}}(k) \\ (k+1) \hat{\mathbf{W}}_{t_{\text{end}}}(k+1) = -\mathbf{Q} \hat{\mathbf{V}}_{t_{\text{end}}}(k) - \mathbf{A}^T \hat{\mathbf{W}}_{t_{\text{end}}}(k) \end{cases} \tag{24}$$

式中, $\hat{\mathbf{V}}_{t_{\text{end}}}(i), \hat{\mathbf{W}}_{t_{\text{end}}}(i)$ 表示矩阵在 t_{end} 处的 i 阶微分变换,其中 0 阶微分变换为:

$$\begin{cases} \hat{\mathbf{V}}_{t_{\text{end}}}(0) = \mathbf{I} \\ \hat{\mathbf{W}}_{t_{\text{end}}}(0) = \mathbf{P} \end{cases} \tag{25}$$

由微分逆变换定义知,矩阵 $\mathbf{V}(t)$ 和 $\mathbf{W}(t)$ 的近似可写成 N 阶泰勒多项式展开形式:

$$\begin{cases} \mathbf{V}(t) = \sum_{k=0}^N \hat{\mathbf{V}}_{t_{\text{end}}}(k) (t-t_{\text{end}})^k \\ \mathbf{W}(t) = \sum_{k=0}^N \hat{\mathbf{W}}_{t_{\text{end}}}(k) (t-t_{\text{end}})^k \end{cases} \tag{26}$$

将式(21)、式(26)代入式(13)中得最优反馈控制输入:

$$\Delta \mathbf{u}(t) = -\mathbf{R}^{-1} \mathbf{B}^T \mathbf{W}(t) \mathbf{V}^{-1}(t) \Delta \mathbf{x}(t) \tag{27}$$

则制导周期 $[t_{\text{now}}, t_{\text{end}}]$ 内,系统的实际控制输入为 $\mathbf{u}(t) = \mathbf{u}_d(t) + \Delta \mathbf{u}(t_{\text{now}})$,其中 $\mathbf{u}_d(t)$ 为标准轨迹对应的控制输入剖面。

综上,微分变换法求解最优反馈控制律的基本流程如图 2 所示。

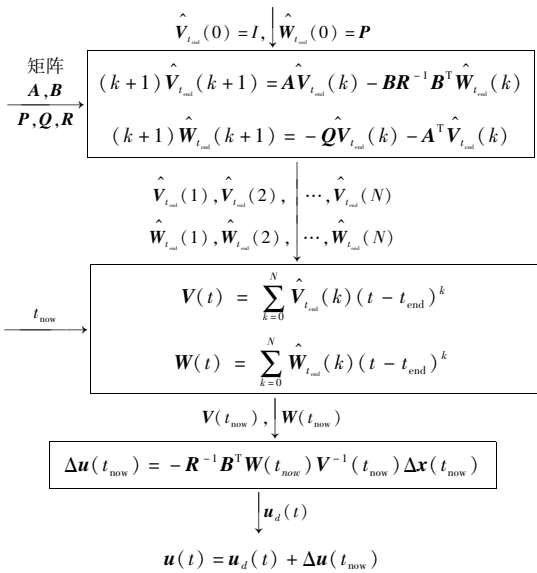


图2 微分变换法求解流程

Fig. 2 Solution flow of the differential transformation

4 仿真分析

以美国 CAV-H 飞行器为例开展仿真分析,其模型参数如表 2 所示^[10],再入飞行任务描述如表 3 所示。

表 2 CAV-H 模型参数

Tab. 2 Model parameters of CAV-H

m/kg	$S_{\text{ref}}/\text{m}^2$	E^*	C_L^*
907	0.483 9	3.24	0.45

表 3 再入飞行任务描述

Tab. 3 Description of the reentry mission

约束条件	约束取值
起始点:	$h_0 = 70 \text{ km}, \theta_0 = 0^\circ, \varphi_0 = 0^\circ,$
边界约束	$V_0 = 6500 \text{ m/s}, \gamma_0 = -1^\circ, \psi_0 = 90^\circ$
终端点:	$h_f \geq 25 \text{ km}, \theta_f = 80^\circ, \varphi_f = 30^\circ$
过程约束	控制量: $\lambda \in [0, 2], \sigma \in [-85^\circ, 85^\circ],$ $n_{\text{max}} = 4g, q_{\text{max}} = 100 \text{ kPa}, Q_{\text{max}} \leq 1700 \text{ kW/m}^2$

由于运动方程非线性强、控制输入灵敏、路径约束严格等因素很大程度上增加了高超声速再入飞行标准轨迹设计的难度,本文选取到达目标点时间最短为性能指标,采用 Radau 伪谱法对再入飞行标准轨迹进行优化设计。Radau 伪谱法的基本原理是在 Legendre-Gauss-Radau 点处同时离散状态和控制变量,利用全局正交插值多项式近似状态和控制变量,并以微分矩阵计算状态变量在离散点处的导数,消除微分方程约束,最终转换为

非线性规划问题。不少文献介绍了伪谱法的具体步骤^[11-12],此处不再赘述。

为降低转换所得非线性规划问题的求解难度,对运动模型作无量纲化处理以增加解的收敛半径:地心距、速度以及时间的无量纲基准分别为 $R_e, \sqrt{R_e g}$ 和 $\sqrt{R_e/g}$ (R_e 为地球半径)。此外,采用串行优化策略,以较少的节点计算满足任务要求的可行轨迹,以此作为设计变量初始猜测值,提高收敛效率。

基于 Radau 伪谱法的标准轨迹优化设计在 MATLAB 环境下基于开源伪谱工具包 GPOPS 进行^[13],配点个数最终取为 70,调用 SNOPT 软件求解离散所得非线性规划问题^[14]。优化所得最短飞行时间为 2095.95 s,相应控制输入、状态变量以及路径约束的变化曲线如图 3~5 中带圈的实线所示。仿真曲线表明飞行器在大气中跳跃式再入,终端经纬度及高度满足终端状态约束;速度大小与方位角变化趋势平缓,航迹倾角呈波浪式变化,导致飞行高度跳跃式降低。再入飞行的大部时间内,飞行器维持最大升阻比状态飞行,意味着射程相同时,最大升阻比飞行对应的飞行时间较短。倾侧角与路径约束均在指定范围内调整,且再入初始阶段热流密度约束起主要影响,后期动压与过载约束起主要作用。

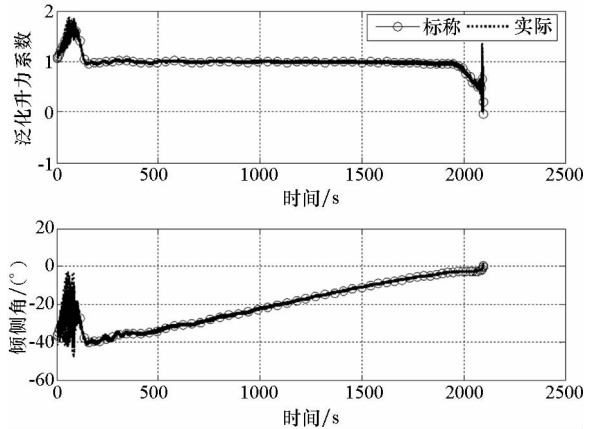


图3 控制输入曲线

Fig. 3 Time histories of control inputs

为验证所研究最优制导策略的有效性,引入如表 4 所示的初始状态偏差、气动参数偏差以及大气模型偏差开展蒙特卡洛 (Monte Carlo) 仿真分析。滚动时域控制性能指标式 (9) 中的权重系数矩阵依据 Bryson 准则选取:

$$Q = \text{diag}([0.03 \quad 3.28 \times 10^7 \quad 6.56 \times 10^7 \dots \\ 0.11 \quad 1.31 \times 10^6 \quad 1.31 \times 10^6]) \\ R = \text{diag}([25 \quad 10.13]) \\ P = 0$$

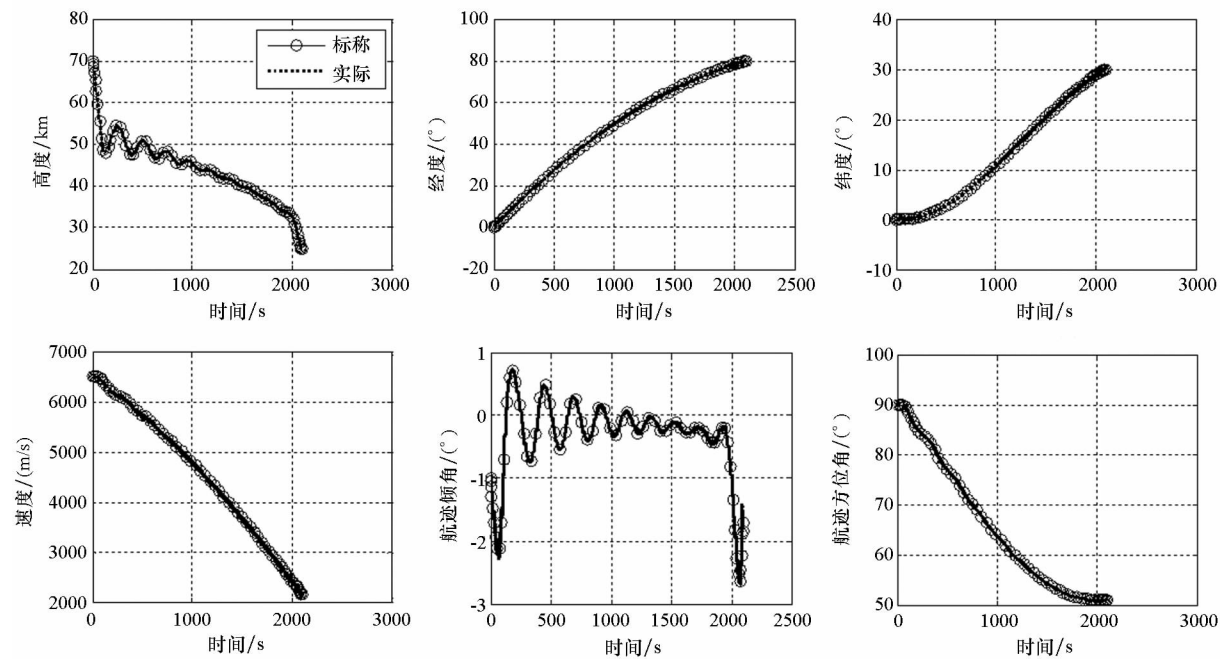


图 4 状态变量曲线
Fig. 4 Time histories of states

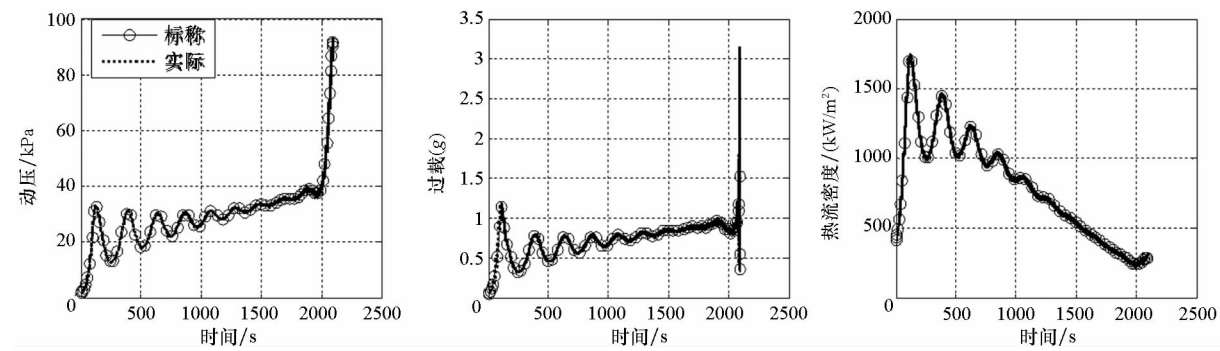


图 5 路径约束曲线
Fig. 5 Time histories of path constraints

Monte Carlo 仿真得到的控制输入、状态与路径约束曲线分别如图 3 ~ 5 中虚线所示。实际再入轨迹在标准轨迹附近小范围内波动,满足模型线性化处理的条件,表明所设计的制导律对于初始状态偏差、气动系数与大气模型不确定性具有较好的鲁棒性。控制变量均在容许范围内调整,且再入初期调整幅度较大,这主要由初始状态偏差造成。

值得注意的是,微分变换法求解两点边值问题得到的实质是最优解的有限阶近似。为对比本文方法性能,在同样仿真配置下,采用文献[4]提出的基于间接伪谱法的最优制导律开展 Monte Carlo 仿真分析,仿真次数取为 100。表 5 给出了两类制导律作用下仿真结果的统计分析,其中 DTM 表示本文方法,而 IPSM 表示基于间接伪谱法的制导律。结果表明,两类制导律均能有效降低外界扰动与模型不确定性造成的影响,且本文方法对终端位置控制精度略优。此外,本文方法计算效率远高于 IPSM:在 2.8 GHz 主频,2 GB 内存的计算机上,采用本文方法进行上述 Monte Carlo 仿真平均耗时 14.103 0 s,而 IPSM 需耗时 37.925 3 s。这主要是由于 IPSM 需进行大量高维矩阵运算。

表 4 偏差因素 3σ 标准分布

Tab. 4 3σ standard dispersions of deviations			
偏差	3σ 取值	偏差	3σ 取值
$\Delta r_0/\text{m}$	± 200	$\Delta \gamma_0/(^{\circ})$	± 0.2
$\Delta \theta_0/(^{\circ})$	± 0.2	$\Delta \psi_0/(^{\circ})$	± 0.2
$\Delta \varphi_0/(^{\circ})$	± 0.2	$C_L, C_D/\%$	± 10
$\Delta V_0/(\text{m/s})$	± 50	$\rho/\%$	± 10

表5 终端状态偏差统计分析
Tab.5 Statistics of terminal dispersion

状态	期望		标准差	
	DTM	IPSM	DTM	IPSM
r/m	0.011 5	-1.555 3	9.919 9	26.745 9
$\theta/(^{\circ})$	-0.024 5	0.024 3	0.220 1	0.275 8
$\varphi/(^{\circ})$	-0.026 7	-0.054 7	0.162 3	0.188 1
$V/(\text{m/s})$	-2.071 3	-0.862 1	13.924 6	13.349 8
$\gamma/(^{\circ})$	-0.100 1	0.112 4	0.014 6	0.123 4
$\psi/(^{\circ})$	-0.002 3	-0.008 3	0.000 9	0.099 5

5 结论

本文研究了基于滚动时域控制的高超声速再入制导问题,提出利用微分变换求解相应两点边值问题以生成最优制导指令的方法。所研究的方法对初始状态偏差、飞行器参数及大气模型等不确定性具有较好的鲁棒性;相对其他最优制导律,本文方法所需计算量小,易于实现。所提出的最优制导律是基于线性化偏差动力学模型推导得出的,未来将研究非线性鲁棒再入制导律;如何直接利用微分变换设计标准轨迹也是下一步的工作。

参考文献 (References)

[1] Joshi A, Sivan K, Amma S S. Predictor-corrector reentry guidance algorithm with path constraints for atmospheric entry vehicles [J]. Journal of Guidance, Control, and Dynamics, 2007, 30(5): 1307-1318.

[2] 徐明亮, 陈克俊, 刘鲁华, 等. 高超声速飞行器准平衡滑翔自适应制导方法[J]. 中国科学: 技术科学, 2012, 42(4): 378-387.

XU Mingliang, CHEN Kejun, LIU Luhua, et al. Quasi-equilibrium glide adaptive guidance for hypersonic vehicles[J]. Science China: Technological Sciences, 2012, 42(4): 378-387. (in Chinese)

[3] Lu P. Regulation about time-varying trajectories: precision entry guidance illustrated [J]. Journal of Guidance, Control,

and Dynamics, 1999, 22(6): 784-790.

[4] Tian B, Zong Q. Optimal guidance for reentry vehicles based on indirect Legendre pseudospectral method [J]. Acta Astronautica, 2011, 68(7/8): 1176-1184.

[5] Peng H J, Gao Q, Wu Z, et al. Optimal guidance based on receding horizon control for low-thrust transfer to libration point orbits [J]. Advances in Space Research, 2013, 51(11): 2093-2111.

[6] Yan H, Ross I M, Alfried K T, et al. Pseudospectral feedback control for three-axis magnetic attitude stabilization in elliptic orbits [J]. Journal of Guidance, Control, and Dynamics, 2007, 30(4): 1107-1115.

[7] Hwang I, Li J H, Du D. Differential transformation and its application to nonlinear optimal control [J]. Journal of Dynamic Systems, Measurement, and Control, 2009, 131(5): 051010-20.

[8] 赵汉元. 飞行器再入动力学与制导[M]. 长沙: 国防科技大学出版社, 1997.

ZHAO Hanyuan. Vehicle reentry dynamics and guidance[M]. Changsha: National University of Defense Technology Press, 1997. (in Chinese)

[9] 阮春荣. 大气中飞行的最优轨迹[M]. 茅振东, 译. 北京: 宇航出版社, 1987.

RUAN Chunrong. Optimal trajectories in atmospheric flight[M]. Translated by MAO Zhendong. Beijing: Astronautics Press, 1987. (in Chinese)

[10] Phillips T H. A common aero vehicle (CAV) model, description, and employment guide[R]. Schafer Corporation for AFRL and AFSPC, 2003.

[11] Ross I M, Karpenko M. A review of pseudospectral optimal control: from theory to flight [J]. Annual Reviews in Control, 2012, 36(2): 182-197.

[12] 杨希祥, 张为华. 基于 Gauss 伪谱法的固体运载火箭上升段轨迹快速优化研究[J]. 宇航学报, 2011, 32(1): 15-21.

YANG Xixiang, ZHANG Weihua. Rapid optimization of ascent trajectory for solid launch vehicles based on Gauss pseudospectral method [J]. Journal of Astronautics, 2011, 32(1): 15-21. (in Chinese)

[13] Rao A V, Benson D A, Darby C, et al. GPOPS: a MATLAB software for solving multiple-phase optimal control problems [J]. ACM Transactions on Mathematical Software, 2010, 37(2): 1-39.

[14] Gill P E, Murray W, Saunders M A. SNOPT: an SQP algorithm for large-scale constrained optimization [J]. SIAM Review, 2005, 47(1): 99-131.

不同系统组合的精密单点定位性能分析*

黄令勇^{1,2,3}, 刘宇玺³, 辛国栋¹, 朱雷鸣¹, 李 五¹, 张 欢¹

(1. 地理信息工程国家重点实验室, 陕西 西安 710054;

2. 信息工程大学 地理空间信息学院, 河南 郑州 450001;

3. 中国天绘卫星中心, 北京 102102)

摘 要:在分析研究星间单差精密单点定位算法和抗差 Kalman 滤波解算模型基础上,利用全球定位系统、全球导航卫星系统、北斗卫星导航系统数据,对单、双、三系统精密单点定位精度和收敛时间进行了分析,得出了以下结论:三系统精密单点定位技术无论定位精度还是收敛速度均最优,多系统组合导航定位有利于提高导航定位精度。

关键词:精密单点定位;三系统;组合定位;北斗卫星导航系统;全球定位系统;全球导航卫星系统

中图分类号:P228 **文献标志码:**A **文章编号:**1001-2486(2017)03-030-06

Performance analysis of different system precise point positioning

HUANG Lingyong^{1,2,3}, LIU Yuxi³, XIN Guodong¹, ZHU Leiming¹, LI Wu¹, ZHANG Huan¹

(1. State Key Laboratory of Geo-information Engineering, Xi'an 710054, China;

2. School of Surveying and Mapping, Information Engineering University, Zhengzhou 450001, China;

3. China Aerospace Surveying and Mapping Satellite Center, Beijing 102102, China)

Abstract: The PPP(precise point positioning) algorithm based on the single differencing between satellites and the robust Kalman filter model was studied and analyzed. And then, the global position system, global navigation satellite system and BeiDou navigation satellite system data had been used to analyze to the positioning precision and convergence time of single, double and three-system PPP. Finally, the conclusion were drawn as follows: the positioning precision or the convergence speed of the three-system PPP technology is optimized, and multiple-system integrated navigation and positioning can improve the precision of navigation and positioning.

Key words: precise point positioning; triple-system; integrated positioning; BDS; GPS; GLONASS

随着我国北斗卫星导航系统(BeiDou navigation satellite System, BDS)的建成运行,精密单点定位(Precise Point Positioning, PPP)技术研究再次成为热点^[1]。由文献[2]可知,单系统PPP依然存在以下问题:①卫星信号遮挡情况下可视卫星数少,容易造成卫星空间几何结构差,进而影响定位精度;②单系统载波模糊度与位置参数、接收机钟差、对流层延迟等误差分离困难,导致收敛时间长。而当全球定位系统(Global Positioning System, GPS)卫星较少时,增加全球导航卫星系统(GLObal NAVigation Satellite System, GLONASS)卫星可有效提高GPS PPP收敛速度和定位精度^[3]。随着我国BDS运行以及GLONASS卫星补网完善,有必要进一步对GPS、GLONASS、BDS组合双系统甚至三系统PPP性能进行分析,

以充分发挥系统组合定位优势。

1 PPP解算模型

首先给出多系统观测模型:

$$\begin{cases} P_i^G = \rho + cdt_r^G - cdt_s^{s,G} + T^G + I_i^G + \varepsilon_{p_i}^G \\ L_i^G = \rho + cdt_r^G - cdt_s^{s,G} + T^G - I_i^G + \lambda_i^G N_i^G + \varepsilon_{\phi_i}^G \end{cases} \quad (1)$$

式中, ρ 为站星距离, c 表示光速, dt_r 和 dt^s 分别为接收机钟差和卫星钟差, T 为对流层延迟, I 为电离层延迟, N 为整周模糊度, λ 为载波波长, ε_p 表示伪距观测噪声, ε_ϕ 表示载波观测噪声,上标G和s分别代表卫星系统和卫星,下标i和r分别表示信号频率和接收机。

为消除电离层延迟,多采用无电离层组合:

* 收稿日期:2015-12-22

基金项目:国家自然科学基金资助项目(41674019);国家重点研发计划资助项目(2016YFB0501701);地理信息工程国家重点实验室开放基金资助项目(SLKGE2015-M-2-1)

作者简介:黄令勇(1987—),男,山东嘉祥人,工程师,博士,E-mail:hlylj87@126.com

$$\begin{cases} P_{\text{IF}}^G = \frac{f_1^2 P_1^G - f_2^2 P_2^G}{f_1^2 - f_2^2} = \rho + cdt_r^G - cdt^{s,G} + T^G + \varepsilon_{P_{\text{IF}}}^G \\ L_{\text{IF}}^G = \frac{f_1^2 L_1^G - f_2^2 L_2^G}{f_1^2 - f_2^2} = \rho + cdt_r^G - cdt^{s,G} + T^G + \lambda_{\text{IF}}^G N_{\text{IF}}^G + \varepsilon_{\phi_{\text{IF}}}^G \end{cases} \quad (2)$$

式中, P_{IF} 与 L_{IF}^G 分别为伪距、载波消除电离层组合。

除需消除电离层误差, PPP 解算之前还需进行数据预处理和相应误差改正, 以及多系统的时空基准统一^[4]。非差数据预处理是影响 PPP 精度的一个重要因素, 其主要内容是进行周跳探测与修复。由于钟跳将导致所有观测卫星上所有频率的伪距、相位观测值产生类似周跳的数据阶跃, 但它与周跳有本质区别。为避免周跳探测误判, 周跳探测之前需进行实时钟跳探测与修复。

$$\begin{cases} \Delta P(j) = P(j) - P(j-1) \\ \Delta L(j) = L(j) - L(j-1) \end{cases}, \text{构造检验量 } S: \begin{cases} S^k(j) = \Delta P^k(j) - \Delta L^k(j) \\ |S^k(j)| > k_1 \approx 0.001c \end{cases} \quad (3)$$

式中, j 表示历元, k 为卫星, k_1 为阈值。

对于某一历元, 当且仅当所有可用卫星满足式(3)时, 才认为该历元时刻可能存在钟跳或所有卫星同时发生大周跳。利用式(4)计算钟跳候选值 m , 并确定实际钟跳值 J_s 。

$$\begin{cases} m = \alpha \cdot \left(\sum_{j=1}^n S^k \right) / (nc) \\ J_s = \begin{cases} \text{INT}(m) & |m - \text{INT}(m)| \leq k_2 \\ 0 & |m - \text{INT}(m)| > k_2 \end{cases} \end{cases} \quad (4)$$

式中: α 为系数因子, $\alpha = 10^3$; INT 为取整函数; k_2 为阈值, $k_2 = 10^{-7} \sim 10^{-5}$ 。

完成上述钟差修复以后, 即可通过常用的宽窄巷 (Melbourne-Wubben, MW) 组合和无几何 (Geometry-Free, GF) 组合进行周跳探测与修复, 具体算法可参见文献[5-6]。

接下来对式(2)线性化并矩阵化。

$$\mathbf{L} = \mathbf{B}\mathbf{X} + \mathbf{\Delta} \quad (5)$$

式中, \mathbf{X} 为待估参数 (主要包括接收机位置、浮点模糊度、接收机钟差、对流层延迟误差), \mathbf{B} 为矩阵系数, \mathbf{L} 和 $\mathbf{\Delta}$ 分别为观测值和随机噪声。

随机模型为:

$$\begin{cases} E(\mathbf{\Delta}) = \mathbf{0} \\ \text{VAR}(\mathbf{\Delta}) = \sigma_0^2 \mathbf{Q} \end{cases} \quad (6)$$

式中, σ_0^2 为单位权方差, \mathbf{Q} 为协因数阵。

对于非差 PPP, 为准确分离接收机钟差与模糊度参数, 必须对伪距观测值合理定权, 而实际定

位中多系统组合定位多依据经验给出各系统间伪距观测值的权, 不够严密。此外, 多系统组合 PPP, 还需估计各导航系统时间差参数。根据文献[2]可知, 星间差分可有效消除接收机钟差, 并由此降低对伪距随机模型准确性的要求。下面给出基于星间差分的 GPS/GLONASS/BDS 三系统进行 PPP 观测方程表达式。

$$\begin{cases} \mathbf{L}_{m \times 1} = \mathbf{A}_{m \times (m+6)} \mathbf{X}_{(m+6) \times 1} + \boldsymbol{\varepsilon}_{m \times 1} \\ \text{Cov}_{\varepsilon_L} \sim \mathbf{N}(0, \mathbf{Q}_{LL}) \end{cases} \quad (7)$$

式中: \mathbf{L} 为 m 维观测值; \mathbf{A} 为观测组合系数; \mathbf{X} 为待估参数, $\mathbf{X} = [x, y, z, dT, dt_{r, \text{GR}}, dt_{r, \text{GC}}, N_{\text{IF}}^s]^T$, 其中 x, y, z 为坐标参数, dT 为天顶对流层湿分量, $dt_{r, \text{GR}}$ 为 GPS 和 GLONASS 时间差, $dt_{r, \text{GC}}$ 为 GPS 和 BDS 时间差, N_{IF}^s 为以 s 卫星为参考星的单差无电离层组合模糊度集合; $\text{Cov}_{\varepsilon_L}$ 为对应随机模型。

精密单点定位中, 由于参数较多, 为提高运算效率、克服高阶矩阵求逆困难, 常采用 Kalman 滤波算法进行解算。Kalman 滤波由状态方程和观测方程组成, 状态方程描述相邻时刻状态转移变化, 观测方程描述对状态进行观测的信息。具体 Kalman 方程可表示为:

$$\begin{cases} \mathbf{X}_k = \boldsymbol{\Phi}_{k, k-1} \mathbf{X}_{k-1} + \mathbf{W}_k \\ \mathbf{L}_k = \mathbf{A}_k \mathbf{X}_k + \mathbf{e}_k \end{cases} \quad (8)$$

式中: \mathbf{X}_k 是系统在 $t(k)$ 时刻的状态向量; $\boldsymbol{\Phi}_{k, k-1}$ 为从 $t(k-1)$ 时刻到 $t(k)$ 时刻系统状态的转移矩阵; \mathbf{W}_k 为系统噪声向量; \mathbf{L}_k 为系统在 $t(k)$ 时刻的观测向量; \mathbf{A}_k 为观测方程的系数阵; \mathbf{e}_k 为观测噪声。

Kalman 滤波假定系统观测值、系统噪声为独立零均值高斯白噪声, 且系统噪声与观测噪声互不相关, 此时有:

$$\begin{cases} E(\mathbf{e}_k) = 0 & \text{VAR}(\mathbf{e}_k) = \sigma^2 & \mathbf{Q}_k = \sigma^2 \mathbf{P}_k^{-1} \\ E(\mathbf{W}_k) = 0 & \text{VAR}(\mathbf{W}_k) = \sigma^2 & \mathbf{Q}_{W_k} = \sigma^2 \mathbf{P}_{W_k}^{-1} \end{cases} \quad (9)$$

式中: \mathbf{W}_k 与 \mathbf{W}_{k-1} , \mathbf{e}_k 与 \mathbf{e}_{k-1} , \mathbf{W}_k 与 \mathbf{e}_k 均为不相关的高斯白噪声。

为减弱粗差对 Kalman 滤波解算的影响, 可以引入抗差 Kalman 算法^[7]。

$$\hat{\mathbf{X}}_k = (\mathbf{A}_k^T \bar{\mathbf{P}}_k \mathbf{A}_k + \alpha_k \mathbf{P}_{\bar{\mathbf{X}}_k})^{-1} (\mathbf{A}_k^T \bar{\mathbf{P}}_k \mathbf{L}_k + \alpha_k \mathbf{P}_{\bar{\mathbf{X}}_k} \bar{\mathbf{X}}_k) \quad (10)$$

式中, $\bar{\mathbf{P}}$ 为抗差等价权矩阵, α_k 为自适应因子。

若观测值存在异常, 减小式(10)中等价权矩阵元素, 以控制观测异常对状态估值的影响; 若状态矩阵模型存在异常, 可以减小自适应因子来控制预测信息异常。

2 数据分析

2.1 实验设计

为分析多系统组合 PPP 性能,实验选用 2013 年第 342~348 年积日 7 天 4 个国际导航卫星系统服务组织(International GNSS Service, IGS)测站利用天宝 TRM 59800.00 系列接收机接收的 GPS/GLONASS/BDS 三系统数据分别对 GPS、BDS、GLONASS 单系统、GPS/GLONASS、GPS/BDS 双系统、GPS/GLONASS/BDS 三系统 PPP 精度、收敛时间进行比较分析。具体站点信息具体见表 1。

表 1 各站点信息
Tab. 1 Station information

编号	站名	位置
1	CUTO	南纬 32°,东经 115.9°
2	GMSD	北纬 30.6°,东经 131.0°
3	JFNG	北纬 30.5°,东经 114.5°
4	REUN	南纬 21.2°,东经 55.6°

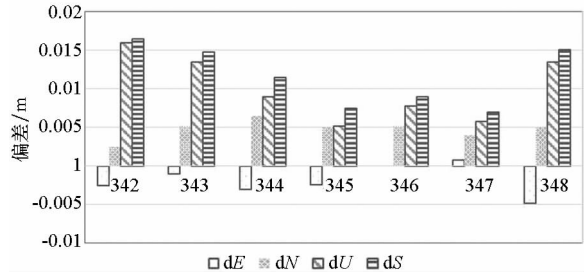
数据处理策略如下:采用伪距/载波无电离层组合观测方程,对相位缠绕、固体潮、相对论进行模型改正;利用欧空局(European Space Agency, ESA)发布的 GPS、GLONASS 精密产品和武汉大学发布的 BDS 精密产品进行卫星钟差和轨道改正;卫星天线采用 IGS_08.atx 绝对天线相位中心模型,接收机进行天线相位中心改正;采用 Saastamoinen 模型改正对流层干分量;采用星间差分消除接收机钟差;将对流层湿分量、坐标、模糊度浮点解作为参数进行估计,采用 Kalman 滤波进行数据处理。

以精密钟差文件中 4 个 IGS 站的日坐标为参考真值,利用 PPP 解算坐标与参考真值在测站坐标系东北天(East North Up, ENU)方向的偏差 dE 、 dN 、 dU 和 PPP 解算坐标与参考真值在测站坐标系中的距离偏差 dS 来分析 PPP 解算精度。定义距离偏差 dS 首次小于 0.1 m 的历元并且该历元以后的 20 历元距离偏差 dS 均未超过 0.1 m 时认为滤波在该历元收敛,从开始定位到完成收敛的时间段称之为收敛时间。由于不同观测时刻、观测条件下 PPP 解算精度和收敛速度不同,为此声明本实验每天 PPP 解算均从接收机独立交换格式(Receiver Independent Exchange Format, RIEF)观测文件记录的 0 时 0 分 0 秒开始。下面以定位精度、收敛时间为评价指标对多系统 PPP 进行相同观测、相同处理策略下的定位性能比较。

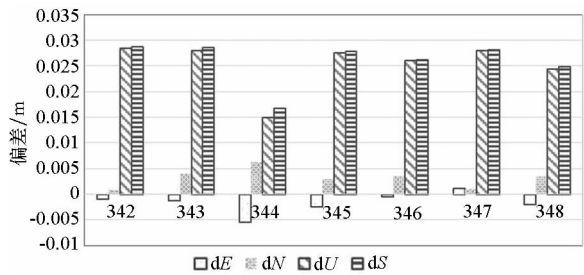
2.2 定位精度分析

2.2.1 单系统静态精度分析

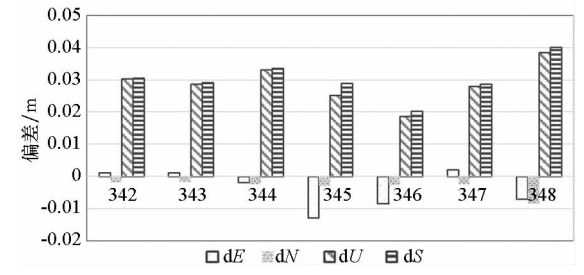
图 1 所示的为 JFNG 站点 7 天单 GPS 静态 PPP 解算结果在 E 、 N 、 U 方向以 PPP 解算位置与参考真值的偏差。由图 1(a)可知,水平方向 dE 和 dN 基本小于 0.5 cm, dU 方向精度较差,但大部分偏差在 1.5 cm 以内,7 天解算结果中最大位置偏差约为 1.6 cm,最小位置偏差约为 0.7 cm,由此可见 GPS 单系统 PPP 静态定位精度较高,能够实现 cm 级。由图 1(b)可知,虽然 GLONASS 单系统 PPP 水平方向偏差在 0.5 cm 以内,但 dU 方向偏差较大,7 天解算结果有 5 天 dU 偏差大于 2.5 cm,由此可见 GLONASS 单定位精度较 GPS 稍差。由图 1(c)可知,BDS PPP 水平方向精度基本小于 1 cm,但 dU 方向偏差明显大于 GPS PPP 和 GLONASS PPP,7 天 dU 平均偏差约为 3 cm,最



(a) GPS 精密单点定位精度分析
(a) Precision analysis of GPS PPP



(b) GLONASS 精密单点定位精度分析
(b) Precision analysis of GLONASS PPP

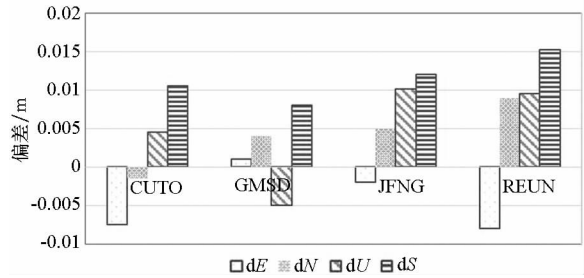


(c) BDS 精密单点定位精度分析
(c) Precision analysis of BDS PPP

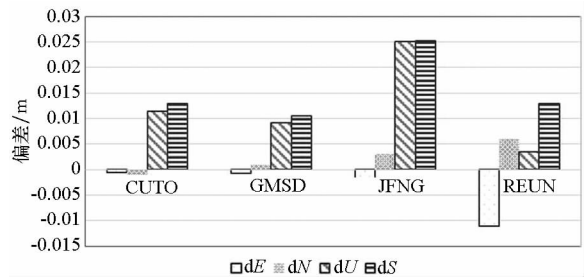
图 1 JFNG 单系统 7 天精密单点定位精度分析
Fig.1 Analysis of JFNG single system 7 days
PPP precision

大偏差接近 4 cm,而 BDS dU 方向偏差过大主要是由于 BDS 尚未发布精确的 BDS 卫星端天线相位中心偏差(Phase Center Offsets, PCO)和天线相位中心漂移(Phase Center Variation, PCV)改正信息等原因。以上三个单系统静态 PPP,水平方向 dE、dU 偏差均优于 dU 方向,主要是与卫星星座在高程方向变化不大有关。由图 1 分析可知,JFNG 站点 7 天静态 GPS PPP 平均定位精度最高,位置偏差约为 1.2 cm;GLONASS 精度次之,约为 2.5 cm;BDS 系统最差约为 3 cm。

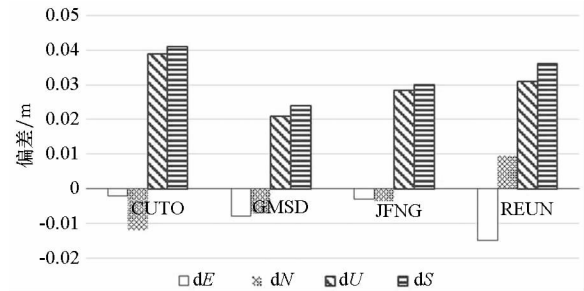
图 2 所示为 4 个站点 7 天 GPS、GLONASS、BDS 单系统静态 PPP 偏差平均值。由图 2 可发现:GPS 精度最高,其距离偏差最大在 1.5 cm 左右,4 站平均位置偏差约为 1.1 cm(如表 2 所示)。GLONASS 精度次之,若除去 JFNG 站点其他 3 站的 7 天平均位置偏差与 GPS 定位结果精度相当,约为 1 cm,但由于 JFNG 站点位置偏差高达 2.5 cm,



(a) GPS 精密单点定位精度分析
(a) Precision analysis of GPS PPP



(b) GLONASS 精密单点定位精度分析
(b) Precision analysis of GLONASS PPP



(c) BDS 精密单点定位精度分析
(c) Precision analysis of BDS PPP

图 2 单系统 7 天平均精密单点定位精度

Fig.2 Single system 7 days PPP average precision

导致 GLONASS PPP 4 站平均位置偏差约为 1.5 cm。以上 4 站 GLONASS 定位结果存在较大差异说明:GPS PPP 解算可靠性高于 GLONASS。BDS PPP 最小位置偏差 dS 均大于 2 cm,4 站 BDS PPP 平均精度约为 3.3 cm。基于分析,有必要加强 BDS 与其他系统的组合定位以提高其精度。

表 2 PPP 精度和收敛时间分析

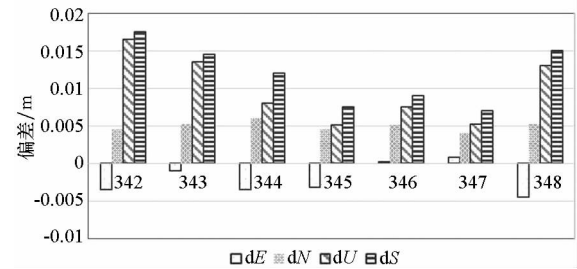
Tab.2 PPP precision and convergence time analysis

	系统		
	GPS	GLONASS	BDS
偏差/cm	1.17	1.55	3.29
收敛时间/min	21	26	110

	系统		
	GPS/ GLONASS	GPS/BDS	GPS/GLONASS/ BDA
偏差/cm	1.13	1.32	1.12
收敛时间/min	16	24	15.6

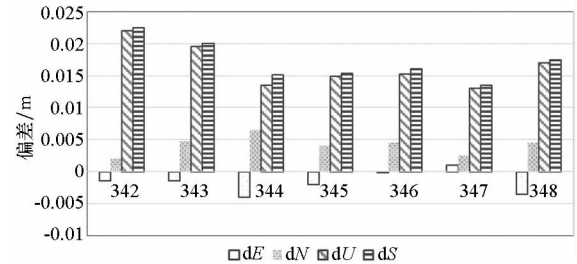
2.2.2 双系统静态精度分析

由图 3 所示的 JFNG 站点 7 天 GPS/GLONASS 和 GPS/BDS 双系统 PPP 精度可知,JFNG 站点 GPS/BDS 组合定位位置偏差基本大于 1.5 cm,而 GPS/GLONASS 组合定位位置偏差除 342 年积日大于 1.5 cm 以外,其他全小于 1.5 cm,甚至有 3 天的定位偏差均小于 1 cm。进一步比较双系统



(a) GPS/GLONASS 双系统定位精度分析

(a) GPS/GLONASS dual-system precision analysis



(b) GPS/BDS 双系统定位精度分析

(b) GPS/BDS dual-system precision analysis

图 3 JFNG 站双系统定位精度分析

Fig.3 Dual-system precision analysis of station JFNG

与三个单系统定位精度, GPS/BDS 显著提高了 BDS PPP 精度, 但 GPS/BDS 组合对 GPS PPP 精度没有改善作用; 而 GPS/GLONASS 双系统组合精度均优于 GPS 和 GLONASS 单系统定位精度。由此发现, GLONASS 与 GPS 组合效果要优于 BDS 与 GPS 组合效果, 这主要是由于目前 BDS 精密产品精度相对 GLONASS 精密产品精度差, 且 BDS 卫星天线相位偏差还未标定改正。

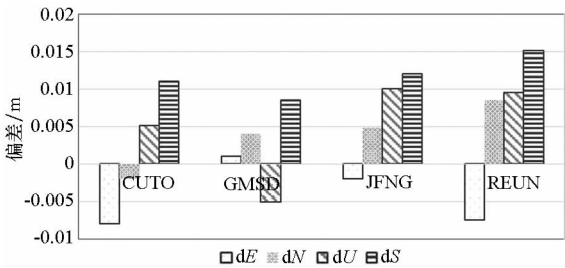
由图 4 所示的 GPS/GLONASS 和 GPS/BDS 双系统 4 站点 7 天平均定位精度分析图可知, REUN 站点基于 GPS/GLONASS 双系统组合解算的 7 天静态 PPP 位置平均偏差值最大, 约为 1.5 cm, 而其他 3 站点平均偏差约为 1 cm。详细比较图 4 中 GPS/GLONASS 各站点定位偏差与图 2 中 GPS、GLONASS 单系统 PPP 4 站点定位偏差可以看出, 双系统组合定位偏差较 GLONASS 单系统定位偏差小, 但定位偏差明显与 GPS 单系统定位偏差趋势相同。以双系统组合 PPP GMSD 站 dU 方向偏差为例, GPS 单系统 dU 偏差为 -0.491 cm, GLONASS 单系统 dU 偏差为 0.8 cm, 而 GPS/GLONASS 组合后 dU 方向偏差为 -0.513 cm, 组合结果与 GPS 单系统结果相差不大, 而与 GLONASS 单系统结果相差甚远, 由此说明双系统组合中 GPS 观测的贡献较 GLONASS 大。这主要是由于双系统组合 PPP 中, 赋予 GPS 观测值的权比较大, 为此结果与 GPS 单系统定位

结果比较接近。GPS/BDS 组合显著改善了 BDS 单系统定位精度, 并且 GPS/BDS 双系统组合后 4 站点不同方向偏差与 GPS 单系统定位偏差具有相同的偏差趋势。同样造成这种现象的主要原因是 BDS 卫星观测质量较 GPS 卫星差, 在组合定位中设置的 GPS 观测值的权大于 BDS 观测值的权, 从而使得双系统组合定位结果与 GPS 单系统解算结果相差不大。

进一步分析表 2 可知, GPS/GLONASS 双系统组合后 4 站静态 PPP 平均位置偏差为 1.13 cm, 较 GPS 单系统平均偏差 1.17 cm 有所改进; 而 GPS/BDS 组合定位平均偏差为 1.32 cm, 其对 GPS 单系统定位精度没有改善作用, 主要是因为静态 PPP 模型强度较大, 低精度 BDS 观测数据对模型强度的影响可以忽略。但双系统组合后, 对 GLONASS 和 BDS 精度改善作用较明显, GPS/GLONASS 将 GLONASS 精度提升了 27%, GPS/BDS 组合将 BDS 精度提升了 60%。

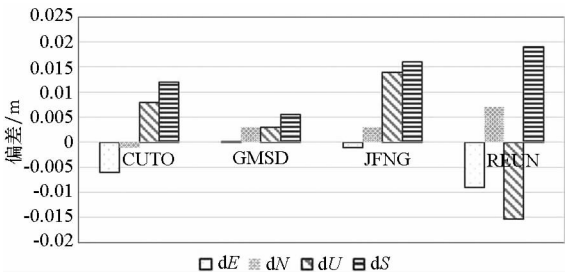
2.2.3 三系统组合精密单点定位精度分析

进一步对三系统精密单点定位精度进行分析, 图 5 所示为 JFNG 站点三系统定位精度分析情况。比较图 5 和图 3 可以发现: JFNG 站点三系统定位精度与 GPS/BDS 组合定位结果相当, 而三系统定位精度反而比 GPS/GLONASS 组合精度差。出现此情况原因是 JFNG 站点位于我国境内, BDS 卫星可观测数目较多, 在三系统组合时 BDS 观测所占权重较大, 进而使得定位精度与 GPS/BDS 系统精度相当, 而低于 GPS/GLONASS 组合定位精度。图 6 所示为三系统定位精度分析情况。比较图 6 与图 4 可以看出: 3 系统组合对其他三站定位精度改进作用较明显, 尤其是 GMSD 站点三系统组合定位后位置偏差由原来 GPS/GLONASS 双系统定位解算的 0.8 cm 减小到 0.5 cm。由此可见, 三系统组合 PPP 可提高 GPS/GLONASS 双系统定位精度, 但三系统定位性能是否有所改进与测站所处的位置以及观测条件有



(a) GPS/GLONASS 双系统定位精度分析

(a) GPS/GLONASS dual-system precision analysis



(b) GPS/BDS 双系统定位精度分析

(b) GPS/BDS dual-system precision analysis

图 4 双系统定位精度分析

Fig. 4 Dual-system precision analysis

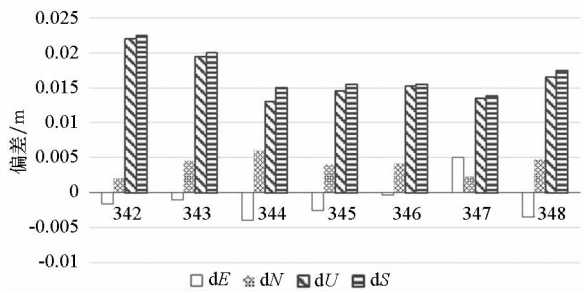


图 5 JFNG 站点三系统定位精度分析

Fig. 5 Triple-system precision analysis of JFNG station

关。根据表 2 可知,4 个站点三系统静态 PPP 平均偏差最小,由此说明在正常观测情况下,三系统静态 PPP 精度最优。

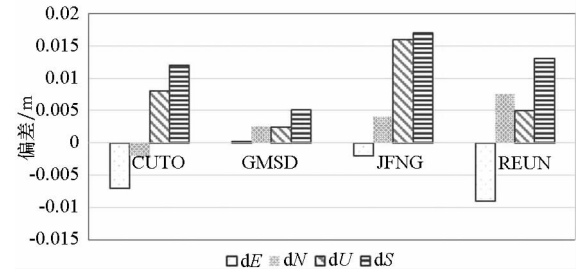


图 6 三系统定位精度分析
Fig. 6 Triple-system precision analysis

2.3 收敛性分析

采用前向 Kalman 滤波进行 PPP 解算,限于篇幅仅给出 4 站 7 天 PPP 平均收敛时间,具体如图 7 所示。

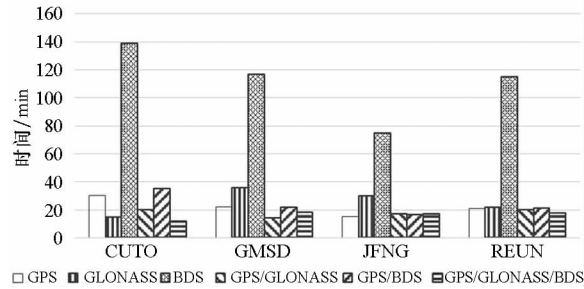


图 7 PPP 收敛时间分析
Fig. 7 PPP convergence time analysis

图 7 所示为 4 个站点 7 天静态 PPP 解算收敛时间的平均值(7 天计算的 PPP 收敛时间进行取平均计算)。从图 7 可以看出,每个站点 BDS 单系统静态 PPP 解算收敛时间均最长,即使收敛时间最短的 JFNG 站点也需要 70 min 才能完成收敛,而其他 3 站均超过了 100 min。4 个站中, GPS、GLONASS 单系统静态 PPP 收敛时间最大不超过 40 min,尤其 JFNG 站点单 GPS 收敛时间仅为 16 min。由此可见,在目前我国 BDS 精密星历钟差以及卫星天线改正信息精度不高或者部分有效信息缺失的情况下,有必要加强 BDS 与 GPS 系统,甚至与 GPS/GLONASS 系统的组合,以加快 BDS 静态 PPP 收敛速度。而由图 7 可知,GPS/BDS 组合后 CUTO 站收敛时间最长,但不超过 40 min,明显比 BDS 单系统收敛时间短;而由其他 3 站可以看出,GPS/BDS 大约在 20 min 即可实现静态 PPP 收敛;同样 GPS/GLOASS/BDS 组合后也显著提升了 BDS 单系统 PPP 收敛速度。GPS/GLONASS 组合除 JFNG 站点对 GPS 收敛时间改

善不明显以外,其他 3 站三系统组合收敛速度均比单 GPS 和单 GLONASS 解算收敛速度快。进一步由图 7 可以看出,GPS/BDS 双系统对 GPS 单系统解算收敛速度的提高效果不明显,甚至组合后收敛速度不如 GPS 单系统收敛速度快。三系统组合收敛速度除 GMSD 站相比 GPS/GLONASS 收敛速度没有提升以外,其他 3 站收敛速度均有所提升,但幅度不大。

进一步统计 4 个站点 7 天静态 PPP 解算的平均收敛时间,具体结果如表 2 所示。由表 2 可知, GPS/GLONASS/BDS 三系统平均收敛时间最短,为 15.6 min;GPS/GLONASS 双系统收敛时间次之,为 16 min。单系统中,GPS 静态收敛速度最快,为 21 min;GLONASS 收敛时间为 26 min;而 BDS 最慢,为 110 min。GPS/BDS 双系统大约对 BDS 收敛速度提高了近 5 倍,由此可见 BDS 与其他系统加强组合定位的必要性。

3 结论

通过对单系统以及 GPS/GLONASS、GPS/BDS 双系统和 GPS/GLONASS/BDS 三系统 PPP 精度和时间收敛的比较发现:三系统组合精密单点定位性能最优;单系统精密单点定位中,GPS、GLONASS PPP 无论精度还是收敛时间均优于 BDS 系统;在组合 PPP 中,BDS 提升组合定位性能作用小于 GLONASS。虽然 BDS 与 GPS 双系统组合难以提升 GPS 单系统定位性能,但 BDS 通过与 GPS 或与 GPS/GLONASS 组合,可明显提高自身定位精度、缩短收敛时间。多系统组合定位可有效提高 PPP 性能。

参考文献 (References)

[1] 施闯,赵齐乐,李敏,等.北斗卫星导航系统的精密定轨与定位研究[J].中国科学:地球科学,2012,42(6):854-861.
SHI Chuang, ZHAO Qile, LI Min, et al. Precise orbit determination of Beidou satellites with precise positioning[J]. Science China: Earth Science, 2012, 42(6): 854-861. (in Chinese)
[2] 张小红,左翔,李盼,等. BDS/GPS 精密单点定位收敛时间与定位精度的比较[J].测绘学报,2015,44(3):250-256.
ZHANG Xiaohong, ZUO Xiang, LI Pan, et al. Convergence time and positioning accuracy comparison between BDS and GPS precise point positioning [J]. Acta Geodaetica et Gartographica Sinica, 2015, 44(3): 250-256. (in Chinese)

GNSS 接收机离散化处理对解扩性能的影响*

刘小汇, 李峥嵘, 陈华明

(国防科技大学 电子科学与工程学院, 湖南 长沙 410073)

摘要:建立扩频信号防混叠滤波、采样、量化与解扩输出关系的数学解析模型,推导得到解扩输出信噪比的解析表达式。分析与仿真表明,当量化位数大于等于 4 bit,解扩得到总的信噪比损失可以分解为由量化引起的损失和滤波加采样引起损失的乘积,且量化器最优限幅系数只与量化位数相关;当量化位数小于 4 bit 时,信噪比损失在一定条件下可近似为量化损失和滤波加采样损失的乘积。当量化位数大于 4 bit、滤波器带宽大于 5 倍码率、采样频率大于 4 倍码率时,再增大上述参数引起的信噪比损失波动小于 0.05 dB,对解扩性能提升不明显。该结论可为实用型全球导航卫星系统接收机前端离散化处理优化设计提供理论指导。

关键词:滤波;采样;量化;信噪比;全球导航卫星系统接收机

中图分类号: TN967.1 **文献标志码:** A **文章编号:** 1001-2486(2017)03-036-05

Discrete processing influence on de-spreading performance of GNSS receiver

LIU Xiaohui, LI Zhengrong, CHEN Huaming

(College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: The analytical model for the output relations of anti-aliasing filtering, sampling, quantization and de-spreading was established. The formulation of de-spreading output SNR (signal to noise ratio) under the influence of three factors was deduced. Analysis and simulation results show that the total SNR loss can be separated to the product of quantization, filtering and sampling loss when the quantization length up to 4 bit; besides, the optimized clipping level coefficient is only related to quantization length. When the quantization length is less than 4 bit, the total SNR loss approximates to the product of the quantization and sampling plus filtering loss. In the condition of the quantization length up to 4 bit, the filter bandwidth up to 5 times of PN (pseudo-random) code rate, the sample rate up to 4 times of PN code rate, increasing the parameters mentioned above cause the SNR loss less than 0.05 dB, thus the de-spreading performance cannot be increased obviously. Results can be used to optimize the design of the front-end of low-cost global navigation satellite system receiver.

Key words: filtering; sampling; quantization; signal to noise ratio; global navigation satellite system receiver

对连续信号离散化是全数字卫星导航接收机工作的前提,在模数转换前,无限带宽的模拟信号经过抗混叠滤波器,将信号带外的噪声滤除,形成有限带宽信号。模数转换由模数转换器完成,包括采样和量化。采样是卫星导航接收机中模拟信号到数字信号转换的第一步,奈奎斯特采样定理指出,当采样频率大于信号带宽的两倍时,由离散的采样点可以无失真地恢复出被采样信号^[1]。卫星导航信号是直接序列扩频信号,对于导航信号的采样,由于需进行精密伪距测量,除了信号不失真外,还要保证解扩后的伪码相关峰不失真。与采样的线性变换不同,量化是典型的非线性变换,量化位数同样决定着输出信号的信噪比,通常认为量化位数越高,信号失真度越小^[1]。

由此可知,卫星导航信号离散化处理的性能是滤波、采样、量化等诸多因素共同影响下的结果。由于量化的非线性使得建立完备的解析式较困难,导致多数对采样和量化的研究立足于数值样本仿真^[2-3]。在解析建模分析方面,为了得到解析结果,多数研究均简化了某些处理环节,导致所得到的结论具有片面性。如文献[4]研究了采样频率对解扩信号不失真的影响,但没有考虑量化的作用,文献[5-6]分析了离散扩频信号经过滤波器和量化器后输出的信噪比损失情况,但忽略了采样频率对输出的影响。

本文通过建立完备的接收机前端处理数学模型,推导扩频信号在滤波、采样和量化下的解析表达式,确立扩频信号解扩输出的信噪比与理想解

* 收稿日期:2016-01-08

基金项目:航天支撑基金资助项目(2011-HTGFKD)

作者简介:刘小汇(1976—),女,广西柳州人,副研究员,博士,E-mail: lululiu_nudt@sina.com

扩信噪比之间的关系,从而得到上述三个因素对解扩性能的影响。

1 信号建模与统计特性分析

1.1 信号建模

如图1所示,卫星导航接收机从天线端接收到射频信号,经过下变频和滤波后得到中频信号:

$$x_{IF}(t) = AD(t)c(t)\cos(2\pi f_0 t + \theta) + w(t) \quad (1)$$

其中: A 为信号幅度; $D(t) = \pm 1$ 为数据信息; f_0 为载波频率; θ 为载波相位; $w(t)$ 为双边功率谱密度等于 $N_0/2$ 的高斯白噪声,服从正态分布; $c(t)$ 为伪随机码,码率为 R_c 。

假设解扩积累均在一个数据位内(不妨设 $D(t)=1$)进行,由于本文只分析滤波器带宽、采样频率、量化位数与解扩输出信噪比的关系,以上三个因素对信号中心频率 f_0 和相位 θ 没有影响,且 f_0 和 θ 可在后续处理中通过相位旋转的方法去除,则式(1)信号部分可以化简成 $s(t) = Ac(t)$ 。

$x_{IF}(t)$ 经过理想匹配滤波器后得到 $x(t)$:

$$x(t) = x_{IF}(t) * h(t) = s_f(t) + w_f(t) \quad (2)$$

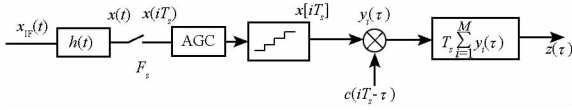


图1 直扩信号离散处理模型

Fig. 1 Direct spread signal model of discrete processing

其中, $s_f(t) = s(t) * h(t)$, $w_f(t) = w(t) * h(t)$ 分别表示信号和噪声经过滤波器的输出。滤波器输出经过周期 T_s 的采样后得到离散信号 $x(iT_s)$,再经过量化器后得到有限精度的离散信号 $x[iT_s]$ 。量化是一种非线性运算,通常采用对称均匀量化^[1],其运算过程可写为:

$$x[iT_s] = Q_B^A[x(iT_s)] \quad (3)$$

量化运算符 Q_B^A 定义为^[7]:

$$Q_B^A(x) = \frac{A}{2} \left[- (2^B - 1) + 2 \sum_{i=-L}^L u\left(\frac{x}{\Delta} - i\right) \right] \quad (4)$$

其中: B 为量化位数;量化级数 $L = 2^{B-1} - 1$; $u(t)$ 为单位阶跃函数;量化间隔 $\Delta = 2V_m/(2^B - 1)$; $V_m = KA$ 为限幅电平,通常取量化信号幅度的倍数, K 为幅度因子。

量化结果与本地伪码的离散形式 $c(iT_s - \tau)$ 进行相关累加,假设积累时间 $T_l = MT_s$ 足够长,则:

$$\begin{aligned} z[\tau] &= T_s \sum_{i=1}^M x[iT_s]c(iT_s - \tau) \\ &= T_s \sum_{i=1}^M \{x_s[iT_s]c(iT_s - \tau) + x_w[iT_s]c(iT_s - \tau)\} \end{aligned} \quad (5)$$

其中: $x_s[iT_s]$ 和 $x_w[iT_s]$ 为量化器输出的信号和噪声部分; τ 表示本地伪码与接收伪码相位之间的延时。为了表达简洁,本文的采样周期、延时、滤波器带宽均为伪码周期 T_c 的归一化函数。

1.2 输出信噪比计算

1.2.1 信号功率

解扩输出的信号部分为:

$$z_s(\tau) = T_s \sum_{i=1}^M x_s[iT_s]c(iT_s - \tau) \quad (6)$$

扩频信号解扩以前,由于信号分量远远小于噪声分量,为了计算方便,可以忽略信号的影响。在信噪比较低时(< -25 dB),得到的近似表达式^[7]为:

$$x_s[iT_s] = s_f(iT_s)K_Q \quad (7)$$

即量化器输出的有用信号为输入的有用信号与量化系数的乘积。设 $Ag = 1/\Delta$,量化系数 K_Q 为^[7]:

$$K_Q = \frac{1}{2\sqrt{2\pi}Ag\sigma_{wf}} \left[2 \sum_{i=-L}^L \exp\left(-\frac{i^2}{2Ag^2\sigma_{wf}^2}\right) \right] \quad (8)$$

式中, σ_{wf}^2 为滤波器输出的噪声项 $w_f(t)$ 的方差。假设采样频率 F_s 满足扩频信号解扩后相关峰不失真的条件时^[4],采样的离散信号解扩可以等效于连续模式 $T_s \rightarrow 0$ 的解扩信号。于是输出信号可进一步计算:

$$z_s(\tau) = AK_Q \int_{-\infty}^{\infty} T_s \sum_{i=1}^M c(iT_s - \tau)c(iT_s - v)h^*(v)dv \quad (9)$$

式中上标“*”为共轭运算符。当 $T_s \rightarrow 0$ 时,有:

$$\lim_{T_s \rightarrow 0} T_s \sum_{i=1}^M c(iT_s - \tau)c(iT_s - v) = T_l \Lambda(\tau - v) \quad (10)$$

其中, $\Lambda(x)$ 为三角波函数,对应的傅里叶变换为 $F_\Lambda(\omega) = \text{sinc}^2(\omega/2)$ 。对于带宽为 b 的理想低通滤波器(b 经码率 R_c 归一化),并且假设滤波器带宽大于伪码码率,即 $b \geq 1$,信号部分为:

$$z_s(\tau) = AT_l K_Q \int_{-b}^b \text{sinc}^2(\pi f) e^{-j2\pi f \tau} df \quad (11)$$

当本地伪码与接收伪码相位严格对齐时($\tau=0$),信号的功率为:

$$\sigma_s^2 = A^2 T_l^2 K_Q^2 \left[\int_{-b}^b \text{sinc}^2(\pi f) df \right]^2 \quad (12)$$

1.2.2 噪声功率

对于噪声部分, 积累解扩输出为:

$$z_w(\tau) = T_s \sum_{i=1}^M x_w[iT_s] c(iT_s - \tau) \quad (13)$$

其中, $x_w[iT_s]$ 为量化器输出的噪声。由前面分析可知, 滤波器输出噪声分量的第 i 个采样点的信号为:

$$w_f(iT_s) = \int_{-\infty}^{\infty} w(iT_s) h^*(v) dv \quad (14)$$

其中, 均值为 $E[w_f(iT_s)] = 0$, 自相关函数为:

$$\begin{aligned} R_{wf}(i, k) &= E[w_f(iT_s) w_f^*(kT_s)] \\ &= E\left[\int w(v) h^*(iT_s - v) dv \cdot \int w(\xi) h(kT_s - \xi) d\xi\right] \end{aligned} \quad (15)$$

其中, $w(t)$ 为高斯白噪声, 其自相关函数为冲击函数:

$$E[w(t_1) w(t_2)] = N_0 \delta(t_1 - t_2) \quad (16)$$

因此式(15)可计算为:

$$\begin{aligned} R_{wf}(i, k) &= N_0 \int_{-\infty}^{\infty} h(iT_s + v) h^*(kT_s + v) dv \\ &= N_0 \int_{-\infty}^{\infty} |H(f)|^2 e^{j2\pi f(k-i)T_s} df \end{aligned} \quad (17)$$

当滤波器为带宽 b 的理想低通滤波器时, 由式(17)得自相关函数为:

$$R_{wf}(i, k) = 2N_0 b \text{sinc}[2\pi(k-i)T_s b] \quad (18)$$

当 $k=i$ 时, 得到滤波器输出的噪声功率为:

$$\sigma_{wf}^2 = 2N_0 b \quad (19)$$

以下求噪声 $w_f(iT_s)$ 经过量化器后的输出。文献[8]指出, 带限白噪声经过量化器后其自相关函数将被改变, 即量化器输出噪声项 $x_w[iT_s]$ 的自相关函数 $R_{xw}(\tau)$, 与输入噪声项的自相关函数 $R_{wf}(\tau)$ 将有所不同。

假设 $f_1(x)$ 与 $f_2(x)$ 是 x 的两种非线性运算过程, 则对于输入的高斯随机过程中的任意两个采样点 ($x_1 = x(t)$ 与 $x_2 = x(t+\tau)$) 而言, 非线性运算输出的互相关函数 $R(\tau)$ 与互相关系数 $\rho(\tau)$ 具有如下的 k 阶偏导数关系^[7]:

$$\begin{aligned} \frac{\partial^k R(\tau)}{\partial \rho(\tau)^k} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{f_1^{(k)}(x_1) f_2^{(k)}(x_2) \exp\left(-\frac{x_1^2 + x_2^2 - 2\rho x_1 x_2}{2(1-\rho^2)}\right)}{2\pi \sqrt{1-\rho^2}} dx_1 dx_2 \end{aligned} \quad (20)$$

其中, $f_i^{(k)}(x)$ 为函数 $f_i(x)$ 对 x 的 k 阶偏导数 ($i=1, 2$)。当非线性运算为量化运算时, 得到 $f_i(x)$ 的表达式为:

$$f_1(x) = f_2(x) = Q_B^A[x] \quad (21)$$

对 x 的一阶偏导数为:

$$\frac{\partial f_1(x)}{\partial x} = \frac{\partial f_2(x)}{\partial x} = \sum_{i=-L}^L u\left(\frac{x}{\Delta} - i\right) \quad (22)$$

由以上推导得到噪声通过量化器后输出的相关函数为:

$$\begin{aligned} R_{xw}(\tau) &= \left(\frac{\Delta}{2}\right)^2 \frac{2}{\pi} \sum_{i=-L}^L \sum_{k=-L}^L \int_0^{\rho_{wf}(\tau)} \frac{\exp\left(-\frac{\Delta^2}{\sigma_{wf}^2} \cdot \frac{i^2 + k^2 - 2ikr}{2(1-r^2)}\right)}{\sqrt{1-r^2}} dr \end{aligned} \quad (23)$$

其中, $\rho_{wf}(\tau)$ 为相关系数, 对于平稳随机过程, 是自相关函数与功率的比值。当 $(i-k)T_s = \tau$ 时, 由式(18)得到输入量化器的噪声相关函数为 $R_{wf}(\tau) = 2N_0 b \text{sinc}(2\pi\tau b)$ 。于是相关系数为:

$$\rho_{wf}(\tau) = \frac{R_{wf}(\tau)}{R_{wf}(0)} = \text{sinc}(2\pi\tau b) \quad (24)$$

将量化器输出的相关函数 $R_{xw}(\tau)$ 对输入功率 σ_{wf}^2 进行归一化, 得到量化归一化相关系数为:

$$\eta_{xw} = \frac{R_{xw}(\tau)}{\sigma_{wf}^2} = \frac{R_{xw}(\tau)}{2N_0 b} \quad (25)$$

得到量化器的输出与本地伪码相乘积累进行解扩, 解扩输出噪声项的自相关函数为:

$$R_{zw}(\tau_1, \tau_2) = T_s T_l \sum_{n=1-M}^{M-1} R_{xw}(nT_s) \Lambda(nT_s + \tau_1 - \tau_2) \quad (26)$$

1.2.3 输出信噪比及损失

扩频信号经过滤波、采样和量化后, 解扩积累后输出的信噪比为:

$$SNR = \frac{R_{zs}(0)}{R_{zw}(0)} = \frac{T_l A^2 K_Q^2 \left[\int_{-b}^b \text{sinc}^2(\pi f) df \right]^2}{T_s \sum_{n=1-M}^{M-1} R_{xw}(nT_s) \Lambda(nT_s)} \quad (27)$$

由于扩频信号的带宽被扩展为 $2R_c$, 定义窄带扩频信号的信噪比为带宽 $2R_c$ 之内的信号与噪声功率的比值, 经过 T_l 的解扩积累后输出的信噪比为:

$$SNR_{\text{ideal}} \approx \frac{T_l A^2}{N_0} \quad (28)$$

定义扩频信号经过滤波、采样和量化后, 信噪比损失为:

$$SNR_{\text{loss}} = \frac{SNR_{\text{ideal}}}{SNR} = \frac{T_s \sum_{n=1-M}^{M-1} \eta_{xw}(nT_s) \Lambda(nT_s)}{K_Q^2 \left[\int_{-b}^b \text{sinc}^2(\pi f) df \right]^2 / 2b} \quad (29)$$

2 仿真分析

2.1 仅有量化时的信噪比损失

忽略滤波器和采样的影响,扩频信号在仅有量化操作时,即 $b \rightarrow \infty, T_s \rightarrow 0$,代入式(29)可得输出的信噪比损失为:

$$SNR_{\text{loss1}} = \frac{\int_{-1}^1 R_{xw}(t) \Lambda(t) dt}{K_Q^2} \approx \frac{R_{xw}(0)}{K_Q^2} \quad (30)$$

如图2所示,量化位数越小,信噪比损失越大,存在使得损失最小的最优限幅因子。另外,当量化位数大于4 bit后,其最小信噪比损失均小于0.05 dB,再增大量化位数对性能影响不大。

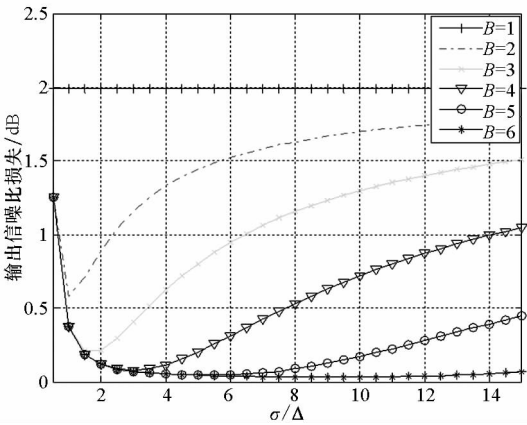


图2 不同量化位数下输出信噪比损失

Fig. 2 SNR_{loss} of different quantization

2.2 仅有滤波和采样时的信噪比损失

由式(29)可知,当只有滤波和采样操作时,信号量化系数 $K_Q = 1$,噪声经过滤波器的相关函数 $R_{xw}(nT_s) = 2N_0 b \rho_{wf}(\tau) = 2N_0 b \text{sinc}(2\pi n T_s)$,于是输出的信噪比损失为:

$$SNR_{\text{loss2}} = \frac{T_s \sum_{n=1-M}^{M-1} \text{sinc}(2\pi n T_s b) \Lambda(n T_s)}{\left[\int_{-b}^b \text{sinc}^2(\pi f) df \right] / 2b} \quad (31)$$

为了保证采样对信号的影响可以等效于连续的三角函数形式,文献[4]指出采样周期应取 $T_s = k/m$ ($k, m \in \mathbb{Z}^+$) 且 k, m 互素。设 $T_s = 1/(2bm)$ ($m > 1$, 且 $2bm$ 非整数),式(31)变为:

$$SNR_{\text{loss2}} = \frac{\sum_{n=1-M}^{M-1} \text{sinc}\left(\frac{\pi n}{m}\right) \Lambda\left(\frac{n}{2mb}\right)}{m \left[\int_{-b}^b \text{sinc}^2(\pi f) df \right]^2} \quad (32)$$

图3所示为 $m \geq 1$ 时(即采样周期满足奈奎斯特采样定理要求),解扩输出信噪比损失与滤

波器带宽的关系。信噪比损失由滤波器带宽 b 和采样周期 T_s 共同影响,即使 T_s 在满足奈奎斯特采样定理的条件下,由于滤波器带宽的影响,输出信噪比仍然有损失,随着带宽的增大,信噪比损失逐渐减小。当滤波器带宽 $b \geq 5$ 时,再增大带宽所带来的输出信噪比只提高不到0.01 dB,性能提升已不明显。

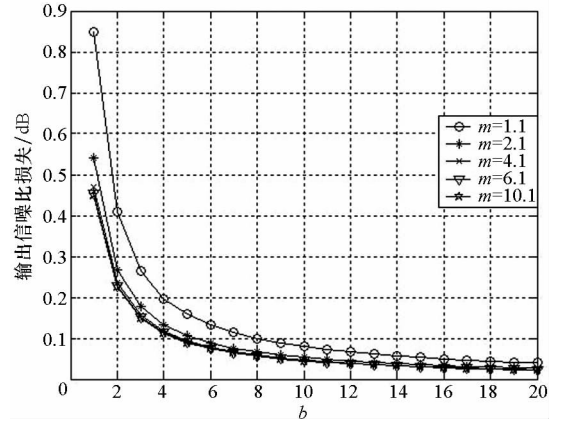


图3 滤波和采样共同作用下的信噪比损失

Fig. 3 SNR_{loss} due to filtering and sampling

2.3 三种因素共同影响下的信噪比

2.3.1 高比特量化

由上述信噪比损失的三个式子,当量化位数 $B \geq 4$ 时, $\eta_{xw}(t) \approx \text{sinc}(\pi t)$,得到如下关系:

$$SNR_{\text{loss}} = SNR_{\text{loss1}} \cdot SNR_{\text{loss2}} \quad (33)$$

2.3.2 低比特量化

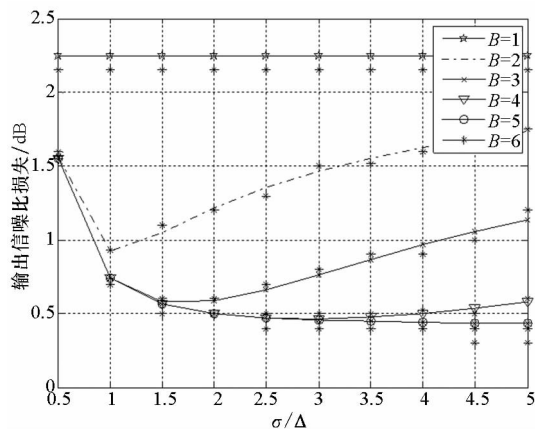
对于量化位数 $B < 4$ 的低比特量化情况,由上节知,量化归一化系数由 b 和 T_s 共同影响:

当 b 不变, T_s 较大时,考虑到式子 $\sum R_{xw}(\tau) \Lambda(\tau)$,由于在一个正负码片区间内 ($\Lambda(\tau)$ 的定义域),有限的采样点使得高、低比特量化下, $R_{xw}(\tau)$ 函数覆盖的面积相差不大,即高、低比特的计算结果相近,因此低比特量化的信噪比损失也可以使用式(33)来近似计算;当采样周期 T_s 较小时,低比特量化的 $R_{xw}(\tau)$ 函数覆盖面积比高比特的小,即实际计算得到的 SNR_{loss} 比使用式(33)的结果偏小。

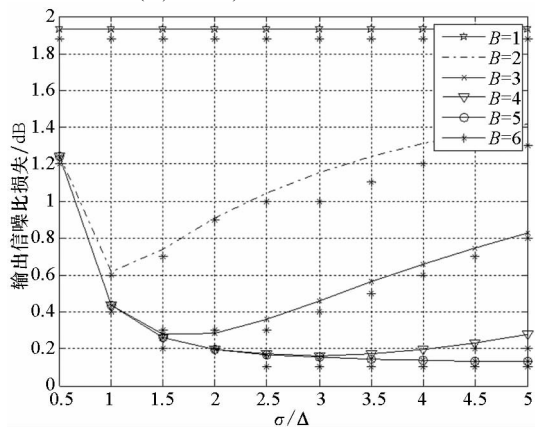
当采样周期 T_s 不变时,随着 b 的增大, $R_{xw}(\tau)$ 在码片区间内的能量减小,高、低比特量化的能量相差不大,低比特可以使用式(33)近似计算;当 b 减小时,低比特量化时 $R_{xw}(\tau)$ 在码片区间内的能量小于高比特的,因此实际计算的信噪比损失比用式(33)的要小。

为了更进一步对比,本文使用了北斗二号导航系统 B1 频点(1561.098 MHz)的民码信号作为

输入信号,输入中频的信噪比 $SNR = -25$ dB,信号带宽 2.046 MHz,使用基带信号处理板进行测试,每种情况测试 10 次。图 4 列出了 $b=2$ 、 $T_s = 1/2 \cdot 200\ 02b$ 与 $b=8$ 、 $T_s = 1/2 \cdot 200\ 02b$ 两种情况下,实际测试数据与理论计算结果的对比情况,其中散点为测试的结果,可见实验结果与理论计算相符。



(a) $b=2, T_s = 1/2 \cdot 200\ 02b$



(b) $b=8, T_s = 1/2 \cdot 200\ 02b$

图 4 测试与理论计算的对比

Fig. 4 Result of testing and theoretic analysis

3 结论

综上所述,全球导航卫星系统接收机的解扩性能,接收前端处理时,由滤波器带宽、采样周期和量化位数共同决定,其输出解扩信号的信噪比与理想解扩信号相比,存在损失:

1) 对于高比特量化(大于等于 4 bit),总的信

噪比损失可以分解为由量化引起的损失与滤波采样引起的损失的乘积;对于低比特量化(小于 4 bit),当采样周期较大(大于 0.1 倍码率)时,或者滤波器带宽较大(大于 4 倍码率)时,总的信噪比损失也可以近似分解为量化与滤波采样两部分损失的乘积,其余情况使用分解形式得到的结果比实际计算结果偏大。

2) 当总的信噪比损失可分解时,最优的限幅系数不随采样周期和前端滤波器带宽而改变。

3) 当量化位数大于 4 bit,滤波器带宽大于 5 倍码率、采样频率大于 4 倍码率后,再增大这些参数,得到的信噪比损失差异小于 0.05 dB,对于解扩性能的提升已不明显。

参考文献 (References)

- [1] Oppenheim A V, Schaffer R W, Buck J R. Discrete-time signal processing [M]. 2nd ed. US: Prentice - Hall, Inc. 1999: 157 - 160.
- [2] Wang Y J, Li M. Novel adaptive method for compensation of timing-skew in time-interleaved ADC [J]. Journal of Systems Engineering and Electronics, 2011, 33(10): 2164 - 2168.
- [3] Santipach W. Signature quantization in fading CDMA with limited feedback [J]. IEEE Transactions on Communications, 2011, 59(2): 569 - 577.
- [4] 许晓勇. 卫星导航接收机高精度建模、分析及优化设计研究[D]. 长沙: 国防科技大学, 2008.
XU Xiaoyong. Study on high-precision modeling, analysis and optimization design for satellite navigation receiver [D]. Changsha: National University of Defense Technology, 2008. (in Chinese)
- [5] Zhao H W, Lian B W, Feng J, et al. Research of 2-bit quantization arithmetic in DS - SS receiver [C]//Proceedings of 4th IEEE Conference on Industrial Electronics and Applications, 2009: 2695 - 2698.
- [6] 王世练, 张尔扬. 直扩数字接收机中 AD 量化比特数的确定[J]. 通信学报, 2004, 25(8): 124 - 128.
WANG Shilian, ZHANG Eryang. Decision of the length of AD quantization in DS digital receiver [J]. Journal of China Institute of Communications, 2004, 25(8): 124 - 128. (in Chinese)
- [7] Borio D. A statistical theory for GNSS signal acquisition [D]. Italy: Politecnico di Torino, 2008.
- [8] Baum R F. The correlation function of Gaussian noise passed through nonlinear devices [J]. IEEE Transactions on Information Theory, 1969, 4(1T-4): 448 - 456.

采用副载波参考波形方法的 GNSS 双载波环多径抑制技术*

徐成涛,唐小妹,黄仰博,陈华明,王飞雪
(国防科技大学 电子科学与工程学院,湖南 长沙 410073)

摘要: 为了实现对高阶二进制偏移载波(BOC)信号的无模糊和抗多径接收,将码相关参考波形的闸波设计思路应用于 GNSS 双载波环路接收方法的副载波锁相环。在副载波锁相环中引入设计的闸波参与信号的相干积分过程,使双载波环法具备抗多径性能且不需要额外引入相关器。对该设计方法的理论和具体实现进行阐述和分析,从副载波多径误差包络和跟踪精度两方面对改进的双载波环路方法性能进行评估。仿真结果显示,采用的算法与双载波环路法相比,可以降低 BOC(1, 1)信号 81.1% 的副载波多径误差包络面积以及 BOC(14, 2)信号 75.1% 的副载波多径误差包络面积。但是,改进的双载波环路法会带来 -6 dB 的相干积分后载噪比损失,降低跟踪精度。因此,在闸波参数设计上,需要谨慎选择以平衡算法的多径抑制和跟踪精度性能。综合来看,该方法适用于解决非弱信号条件下及多径环境下的高阶 BOC 信号接收问题。

关键词: 高阶二进制偏移载波信号;双环路;双载波环路;副载波多径误差包络;载波多径
中图分类号: TN95 **文献标志码:** A **文章编号:** 1001-2486(2017)03-041-06

Multipath mitigation technique of GNSS double phase estimator using subcarrier reference waveform method

XU Chengtao, TANG Xiaomei, HUANG Yangbo, CHEN Huaming, WANG Feixue
(College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: In order to achieve the unambiguous and anti-multipath reception of the BOC (binary offset carrier) signal, the DPE (double phase estimator) was modified by introducing a strobe waveform in the prompt signal correlation process of the SPL (subcarrier phase lock loop) integration. The modified DPE possesses multipath error mitigation performance and employs no additional correlator. The theory and the realization of the proposed approach were explained and analyzed, and the performance of the modified DPE was characterized according to the SMEE (subcarrier multipath error envelope) and the tracking jitter. Simulation results show that, compared with the conventional DPE, the proposed algorithm can provide a reduction in the SMEE area of 81.1% for signal BOC(1,1) and 75.1% for signal BOC(14,2). However, the modified DPE experiences a loss of -6 dB in terms of the post-coherent signal-to-noise ratio, which impacts its tracking precision. Thus, the selection of waveform parameters involves a trade-off between the tracking performances obtained under multipath and thermal noise conditions. Above all, the proposed method is applicable to the receiving problems of multipath environment or non weak signal case.

Key words: high-order binary offset carrier signal; double estimator; double phase estimator; subcarrier multipath error envelope; phase multipath

二进制偏移载波(Binary Offset Carrier, BOC)类信号是在现代化的卫星导航系统设计中引入的新型信号,其频谱具有裂谱特性,且 BOC 信号相比二进制相移键控(Binary Phase Shift Keying, BPSK)信号具有更大 Gabor 带宽,可提高信号伪距测量精度。但是 BOC 信号的自相关函数存在多个峰值点,峰值点的数目和幅度随着 BOC 信号阶数的增加而增大。对于传统的跟踪结构,BOC 信号的自相关特性会导致跟踪环路错

锁现象的发生,同时恶化其对伪距中多径误差的抑制性能^[1]。因此在 BOC 信号接收机设计中需要考虑无模糊接收和高精度跟踪两个关键性能。

目前,针对 BOC 信号的错锁问题已有许多研究^[2-5]。其中,双环路法(Double Estimator, DE)和双载波环路法(Double Phase Estimator, DPE)采用了较为简单的接收结构,并能实现各阶 BOC 信号的无模糊跟踪过程,因此具有较高的研究价值。双环路法采用互相独立的码跟踪锁定环和副

* 收稿日期:2016-01-01
基金项目:国家自然科学基金资助项目(41272385);新世纪优秀人才支持计划资助项目(NECT-11-0317)
作者简介:徐成涛(1987—),男,湖南长沙人,博士研究生,E-mail:xct_nudt@163.com;
王飞雪(通信作者),男,教授,博士,博士生导师,E-mail:wangfeixue365@sina.com

载波跟踪锁定环 (Subcarrier Lock Loop, SLL) 分别跟踪伪随机码和副载波。而双载波环路法在带限条件下将副载波近似为单载波, 并采用副载波锁相环 (Subcarrier Phase Lock Loop, SPLL) 代替副载波跟踪锁定环^[6-7]。对于带限接收机来说, 副载波由于前端滤波的影响不再是方波而是类似于单载波, 因此该方法对副载波的近似更贴近实际情况, 在前端带宽受限的条件下可以提升双环路法的跟踪精度。

双环路和双载波环路的处理算法对多径抑制问题考虑较少。双环路法通过在副载波环中采用窄相关器, 可以一定程度上抑制副载波环路中的多径干扰。双载波环路法由于采用了相位锁定环, 因此没有多径抑制机制^[6-7]。相比于双环路法, 双载波环路法更加需要多径抑制算法来提升其性能, 除了文献[7], 目前尚未有相关的研究发表。

关于多径抑制技术的研究非常丰富^[1, 8-10]。文献[11]介绍了一种载波相位多径抑制加窗相关器, 在实测中可以得到近 20% 的多径误差减轻。在此基础上本文对双载波环路法多径抑制方法进行了研究。

1 多径抑制方法

不考虑载波跟踪的影响, 在接收端的基带 BOC 信号连续表达式如式(1)所示。

$$s(t) = \sum_{i=0}^L Aa_iX(t - \tau_i)D(t - \tau_i)\cos(\varphi_0 + \Delta\varphi_i) + n(t)$$

(1)

其中, 信号经过了 L 条路径到达接收机, $i=0$ 为直达信号, A 为直达信号幅度, $\tau_0=0$ 和 $\varphi_0=0$ 分别为直达信号到达接收机时的时间延迟和载波初相, 而 a_i, τ_i 和 $\Delta\varphi_i$ 分别是第 i 路多径信号相对于直达信号的幅度、时延和载波相位变化量, 其中 $a_0=0$ 。 $D(t)$ 为电文数据, $n(t)$ 为噪声信号。对多径信号模型的研究, 可以转化表示为对信号 $a_i, \tau_i, \varphi_i, L$ 这些参数的研究。 $X(t)$ 为接收信号基带波形, 可以表示成码速率为 f_c 的伪随机码信号 $c(t)$ 与副载波速率为 f_{sc} 的副载波 $sc(t)$ 的乘积。对于 BOC(m, n) 信号有:

$$\begin{cases} sc(t) = \text{sign}[\sin(2\pi f_{sc}t)] \\ f_{sc} = mf_0 \\ f_c = nf_0 \\ f_0 = 1.023 \text{ MHz} \end{cases}$$

(2)

这里 m, n 分别表示副载波和码频率关于 1.023 MHz 的整数倍。双载波环路法中, 在副载波相位跟踪环中使用了单载波信号 $e^{j(\tilde{\omega}_{sc}t + \tilde{\varphi}_{sc})}$ 对副载波进行跟踪 ($\tilde{\omega}_{sc}$ 为估计的副载波频率, $\tilde{\varphi}_{sc}$ 为估计的副载波相位)。在改进方法中, 接收机内部额外采用了一个闸波信号 $w(t)$ 用于相关积分过程。该方法的接收机结构原理如图 1 所示。其中, 伪码由一个标准的延迟锁定环完成跟踪, 载波

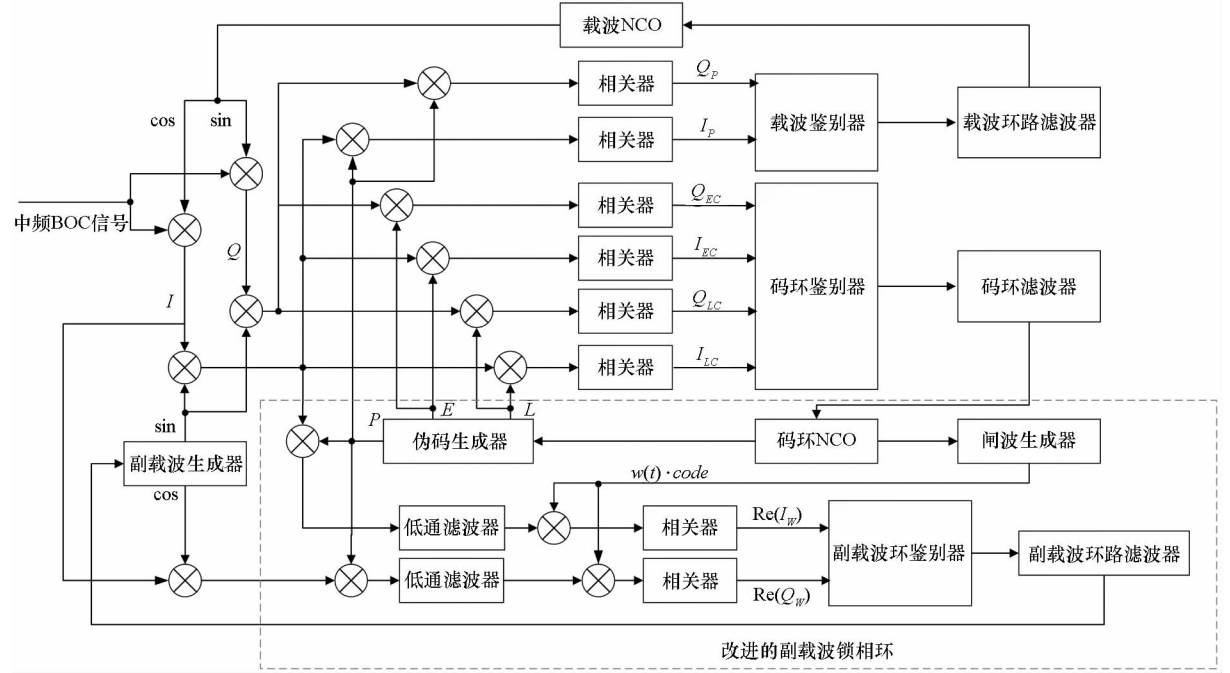


图 1 改进的双载波环路法结构原理图

Fig. 1 Schematic representation of the modified DPE

则由相位锁定环完成跟踪,而副载波的延迟由改进后的相位锁定环(modified SPL)进行恢复。在该副载波锁相环中,在伪码相关之后放置了一个低通滤波器用于消除相关带来的副载波二倍频项。 I_p 和 Q_p 是本地即时码和即时副载波与输入信号相干积分后同步和正交支路结果, I_{EC} 和 Q_{EC} 是本地超前码和即时副载波与输入信号相干积分后同步和正交支路结果, I_{LC} 和 Q_{LC} 是本地滞后码和即时副载波与输入信号相干积分后同步和正交支路结果。

设计闸波如图2所示,其表达式为:

$$\begin{cases} w(t) = \sum_{i=0}^{\infty} g(t - iT_c) c_i(t) \\ g(t) = \sum_{j=1}^{m+n} \omega_j p(t - j\mu) \end{cases} \quad (3)$$

这里 $g(t)$ 为基本闸波,可以看作由 $m+n$ 个方波 $p(t)$ 组成。闸波信号 $w(t)$ 出现在每个扩频码的边沿处,其中 m 为出现在扩频码边沿前的方波个数, n 为出现在边沿后的方波数目。 ω_j 为第 j 个方波的幅度, $c_i(t)$ 是第 i 个扩频码片, T_c 是码片宽度, μ 是闸宽,即 $p(t)$ 宽度。假设总积分长度为 T_{coh} ,则改进的副载波锁相环的 I/Q 支路相干积分结果 I_w 和 Q_w 可以表示为 $s(t)e^{j(\omega_{sc}t + \tilde{\varphi}_{sc})}$ 和 $w(t - \varepsilon)$ 的相关函数,其中 ε 为码延迟估计误差。双载波环接收机的环路鉴别器输入如式(4)所示:

$$\begin{cases} I_w(\varepsilon) = \frac{1}{T_{coh}} \int_0^{T_{coh}} s(t) \sin(\tilde{\omega}_{sc}t + \tilde{\varphi}_{sc}) w(t - \varepsilon) dt \\ = \sum_{i=0}^L P_{IW_i} + N_{IW} \\ P_{IW_i} \approx Aa_i D_k R_w(\varepsilon - \tau_i) \text{sinc}\left(\frac{\omega_{es} T_{coh}}{2}\right) \times \\ \cos\left(\varphi_{es} + \Delta\varphi_{si} + \frac{\omega_{es} T_{coh}}{2}\right) e^{j\Delta\varphi_i} \\ Q_w(\varepsilon) = \frac{1}{T_{coh}} \int_0^{T_{coh}} s(t) \cos(\tilde{\omega}_{sc}t + \tilde{\varphi}_{sc}) w(t - \varepsilon) dt \\ = \sum_{i=0}^L P_{QW_i} + N_{QW} \\ P_{QW_i} \approx Aa_i D_k R_w(\varepsilon - \tau_i) \text{sinc}\left(\frac{\omega_{es} T_{coh}}{2}\right) \times \\ \sin\left(\varphi_{es} + \Delta\varphi_{si} + \frac{\omega_{es} T_{coh}}{2}\right) e^{j\Delta\varphi_i} \end{cases} \quad (4)$$

其中, $R_w(\cdot)$ 为 $w(t)$ 和扩频码 $c(t)$ 之间的相关函数, D_k 为第 k 次相关的电文比特。本地副载波与信号中副载波的频率差和相位差分别为 ω_{es} 和 φ_{es} 。 $\Delta\varphi_{si}$ 为本地载波和信号载波的相位差。 I_w 和 Q_w 中两个独立的零均值高斯随机噪声分别为 N_{IW} 和 N_{QW} ,它们的功率谱密度为 $\frac{N_0}{T_{coh}} \frac{2(m+n)\mu}{T_c}$ 。对于传

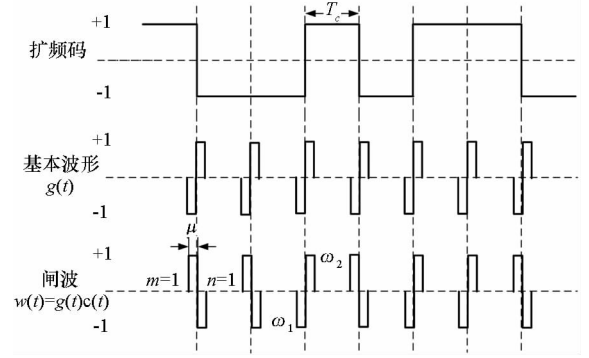


图2 闸波 $w(t)$ 波形示意图

Fig.2 Schematic diagram of strobe waveform $w(t)$

统的双载波环路法,参与相关的不是 $w(t - \varepsilon)$ 而是 $c(t - \varepsilon)$,此时 $R_w(\cdot)$ 可用BPSK码自相关函数 $R_{BPSK}(\cdot)$ 代替。

推导可得 R_w 在分段点 $k\mu$ 处的取值表示为:

$$R_w(k\mu) = \begin{cases} 0, & k \leq -LW - m \\ \frac{1}{LW} \sum_{j=1}^{LW+m-k} \omega_j, & -LW < k \leq -LW + n \\ \frac{1}{LW} \sum_{j=1}^{m+n} \omega_j, & -LW + n < k \leq -m \\ \frac{1}{LW} \sum_{j=m+k+1}^{m+n} \omega_j, & -m < k < n \\ 0, & k \geq n \end{cases} \quad (5)$$

R_w 在这些分段点之间呈线性关系。 $LW = T_c/\mu$ 为一个码片内的最大方波数目。副载波锁相环的鉴别器可采用 \arctan 相干鉴别器,即:

$$\delta_{cp} = \arctan\left(\frac{\text{Re}\{Q_w(\varepsilon)\}}{\text{Re}\{I_w(\varepsilon)\}}\right) \quad (6)$$

为了单独研究副载波环路的跟踪情况,假设本地扩频码可以准确同步,且多径数量 $L=1$,则副载波信号的多径误差可表达为:

$$\varepsilon_{MP} = \arctan\left(\frac{a_1 R_w(\varepsilon - \tau_1) \sin\Delta\varphi_{s1} \cos\Delta\varphi_1}{R_w(\varepsilon) + a_1 R_w(\varepsilon - \tau_1) \cos\Delta\varphi_{s1} \cos\Delta\varphi_1}\right) \quad (7)$$

根据式(7)可知,副载波环路的多径误差主要来自于相关函数 $R_w(\tau_1)$ 。若当 $\tau > \tau_c$ 时, $R_w(\tau) = 0$,则延迟大于 τ_c 的多径信号的多径误差受到该副载波锁相环的抑制作用。对于传统的双载波环路法, $R_{BPSK}(\tau_1)$ 在正负一个码片以内都大于0,因此总是受到延迟小于一个码片的多径信号的影响。通过对 $R_w(\cdot)$ 表达式中非零区域的限制,可以改善其多径抑制性能。

对闸波的设计准则可以归纳为:

1) 当伪码延迟误差 $\varepsilon=0$ 时,鉴别器的输出结果应达到一个最大值,以保证正常跟踪信号时尽量减少信号能量损失。

2) 相关函数 R_w 的非零区域应该尽量小, 以得到较好的多径误差抑制效果。

3) $w(t)$ 的宽度不能超过一个码片。

鉴于此, 闸波的设计准则可以由式(8)表达。

$$\begin{cases} \sum_{j=m+1}^{m+n} \omega_j > 0 \\ \sum_{j=1}^{m+n} \omega_j = 0 \\ (m+n)\mu < T_c \end{cases} \quad (8)$$

根据上述设计准则, 选取的设计参数为 $m=1, n=1, \omega_1=-1, \omega_2=1, \mu=T_c/4$ 。对应的相关函数 R_w 如图 3 所示。

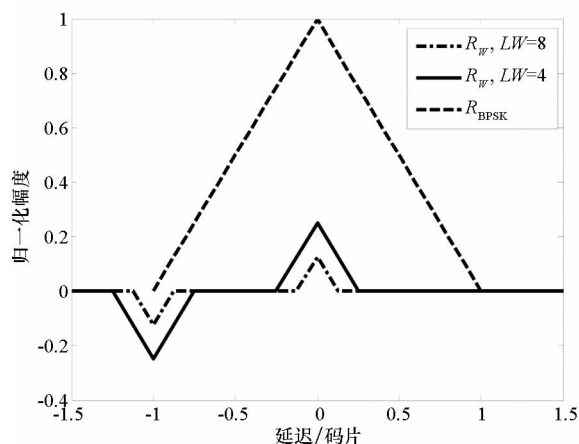


图 3 $w(t)$ 与伪码的相关函数曲线

Fig. 3 Cross correlation function of referenced waveform $w(t)$ and pseudorandom code

改进的双载波环路法的多径误差在式(9)中给出, 与传统双载波环路法的分析不同的是, 其中并未忽略伪码相关函数的影响。

$$\tau_{MP} = \frac{1}{2\pi f_{sc}} \arctan \left[\frac{a_1 \left(1 - \frac{\tau_1}{\eta \mu T_c} \right) \sin(2\pi f_{sc} \tau_1) \cos \Delta \varphi_1}{1 + a_1 \sin(2\pi f_{sc} \tau_1) \cos \Delta \varphi_1} \right], \quad \tau_1 > \eta \mu T_c \quad (9)$$

改进的双载波环路法的副载波多径误差包络 (Subcarrier Multipath Error Envelop, SMEE) 为所有多径载波相位偏移 $\Delta \varphi_1$ 下, 副载波多径误差的最大值和最小值。得到改进的双载波环路法的副载波多径误差包络为:

$$\tau_{MP} = \frac{1}{2\pi f_{sc}} \arctan \left[\frac{a_1 \left(1 - \frac{\tau_1}{\eta \mu T_c} \right) |\sin(2\pi f_{sc} \tau_1)|}{1 + a_1 |\sin(2\pi f_{sc} \tau_1)| \cos(2\pi f_{sc} \tau_1)} \right] \quad (10)$$

$$\tau_{MP} = -\frac{1}{2\pi f_{sc}} \arctan \left[\frac{a_1 \left(1 - \frac{\tau_1}{\eta \mu T_c} \right) |\sin(2\pi f_{sc} \tau_1)|}{1 - a_1 |\sin(2\pi f_{sc} \tau_1)| \cos(2\pi f_{sc} \tau_1)} \right] \quad (11)$$

当多径延迟 $\tau_1 \geq \eta \mu T_c$ 时, $R_w(\varepsilon)$ 等于零, 此时鉴别器输出为 φ_{es} , 即副载波锁相环可以正确估计副载波相位的偏差。

2 性能仿真

从副载波多径误差包络和跟踪精度两个方面对改进的双载波环路法的性能进行分析。

2.1 多径抑制性能

对双环路法、双载波环路法和本文改进的双载波环路法的副载波多径误差包络进行仿真分析, 采用 BOC(1, 1) 和 BOC(14, 2) 信号分别作为低阶 BOC 和高阶 BOC 信号的代表, 其在无限带宽条件下的仿真结果如图 4 和图 5 所示。另外, 对于双环路法, 计算过程中同样考虑了伪码相关函数在副载波延迟锁定环鉴别器中的影响。仿真中, 多径数量为 1, 多径信号相对直达信号的幅度衰减为 3 dB。

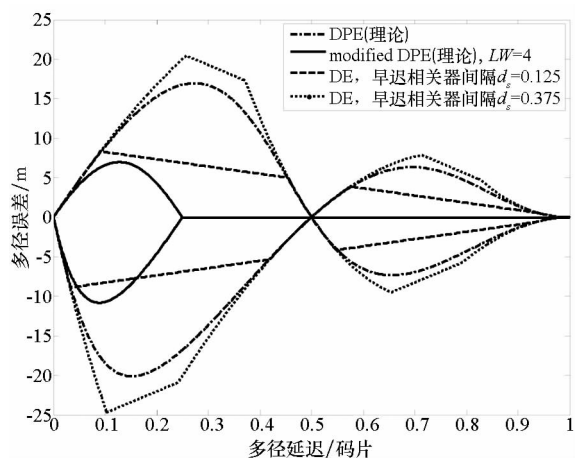


图 4 BOC(1, 1) 信号的副载波多径误差包络曲线仿真
Fig. 4 BOC(1, 1) SMEE evaluated in the absence of front-end filtering

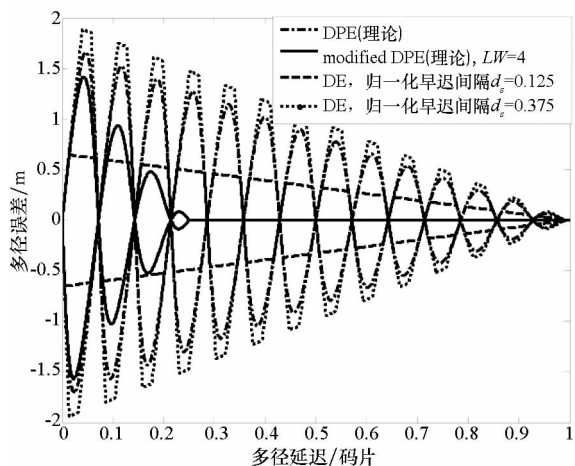


图 5 BOC(14, 2) 信号的副载波多径误差包络曲线仿真
Fig. 5 BOC(14, 2) SMEE evaluated in the absence of front-end filtering

由图4和图5的结果可以得到,改进的双载波环路法相比于双环路法和双载波环路法,拥有最小的副载波多径误差包络,即更强的多径抑制能力。于BOC(1,1)信号而言,改进的双载波环路法的副载波多径误差包络面积相比双载波环路法的减少了81.1%;于BOC(14,2)信号,改进的双载波环路法的副载波多径误差包络面积相比双载波环路法的减少了75.1%。相比于早迟相关器间隔 $d_s = 0.125$ 的双环路法,改进的双载波环路法的副载波多径误差包络面积分别减少了64.4%和53.2%。

可以看到,改进的双载波环路法在BOC(1,1)信号和BOC(14,2)信号上的多径误差抑制性能有所区别。主要原因是两个信号的码率与副载波频率比例不同,BOC(1,1)信号的频率比为1:1,BOC(14,2)信号的频率比为1:7。仿真中改进的双载波环路法的设计参数可抑制延迟大于 $0.25T_c$ 的多径信号,对于BOC(1,1)信号和BOC(14,2)信号而言, $0.25T_c$ 分别等价于 $0.25T_{sc}$ 和 $1.75T_{sc}$ 。因此,在改进的双载波环路法上,BOC(1,1)信号和BOC(14,2)信号的副载波多径误差包络的长度不同。

减小 μ 的数值可降低多径误差包络面积,但是也会带来跟踪性能的下降。因此在设计改进的双载波环路法的参数时需要对多径抑制性能和跟踪精度性能进行折中考虑。

2.2 热噪声性能

参照载波相位锁定环的热噪声性能分析副载波锁相环的热噪声性能,结果如式(12)所示:

$$\left\{ \begin{array}{l} \sigma_\tau \approx \frac{1}{2\pi f_c} \sqrt{\frac{B_{\text{SPLL}}}{\text{CNR}_{\text{eq}}} \left(1 + \frac{1}{2T_{\text{coh}} \text{CNR}_{\text{eq}}} \right)} \\ \text{CNR}_{\text{eq}} = L_c \text{CNR} \left(\frac{s_1^2}{2} \right) \end{array} \right. \quad (12)$$

式中: B_{SPLL} 为副载波锁相环的环路带宽; s_1^2 为副载波近似为单载波带来的能量损失,即双载波环路法相对于双环路法的能量损失,其大小取决于接收机前端带宽,在最坏条件下为-0.91 dB。由于改进的双载波环路法在一个码片内的积分长度从原来的 T_c 降低为 $(m+n)\mu$,其副载波锁相环相干积分后的载噪比CNR相比双载波环路法的额外产生了 L_c 大小的能量损失:

$$L_c = 1 / \left(\frac{2(m+n)\mu}{T_c} \right) \quad (13)$$

从式(13)可以看出,载噪比的损失随着 μ 的降低而增加。对于本文 $m=n=1, LW=4$ 的条件而

言,改进的双载波环路法相比原双载波环路法的载噪比损失为 $L_c = -6$ dB。如果同时扩大相干积分总长度 T_{coh} ,可以在一定程度上弥补这一损失。

对BOC(1,1)和BOC(14,2)信号使用改进的双载波环路法和双环路法在不同载噪比下的跟踪性能进行仿真,并选取了环路带宽 $B_{\text{SPLL}}/\text{SLL}$ 分别为1 Hz、2 Hz和5 Hz的条件,结果如图6和图7所示。BOC(1,1)和BOC(14,2)信号采用了不同的前端接收带宽,分别是8 MHz和32 MHz。双环路法的相关器间隔为0.125码片。

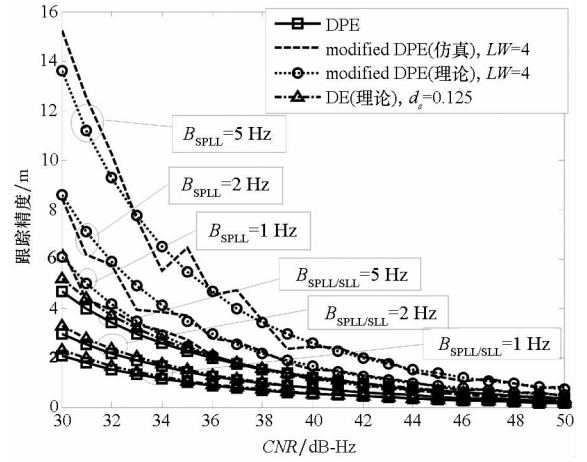


图6 改进的双载波环路法与传统的双载波环路法在BOC(1,1)信号下的跟踪误差仿真结果 (双边带宽为8 MHz)

Fig. 6 Tracking error variance comparison between modified DPE and DPE for BOC(1,1) (double-side bandwidth is 8 MHz)

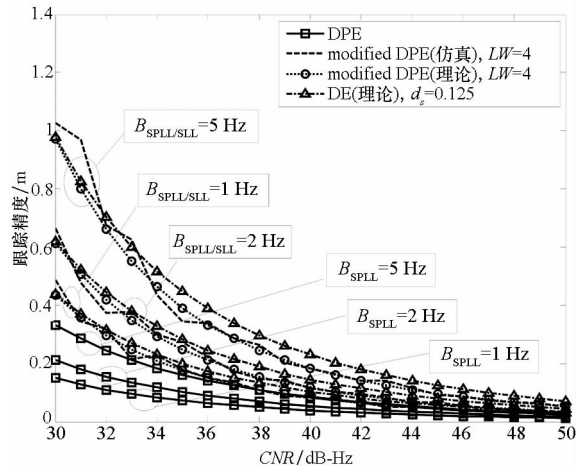


图7 改进的双载波环路法与传统的双载波环路法在BOC(14,2)信号下的跟踪误差仿真结果 (双边带宽为32 MHz)

Fig. 7 Tracking error variance comparison between modified DPE and DPE for BOC(14,2) (double-side bandwidth is 32 MHz)

同样的信号在不同环路带宽的条件下,带宽越小,跟踪性能越好。BOC(14, 2)信号相比 BOC(1, 1)信号具有更高的跟踪精度,将仿真结果与载噪比降低 6 dB 时的理论曲线进行对比,结果显示二者具有较好的一致性,验证了前述分析过程的正确性。此外,BOC(14, 2)信号下改进的双载波环路法与双环路法的跟踪精度相近,BOC(1, 1)信号下改进的双载波环路法与双环路法精度相差约 6 dB(载噪比),这主要是由前端带宽与副载波频率的相对关系不同导致的,说明改进的双载波环路法在前端带宽相对较窄的条件下跟踪精度更具优势。

3 结论

本文在双载波环路法的基础上提出了一种改进的双载波环路法,用于具有无模糊和抗多径性能的 BOC 信号的跟踪。算法采用了经过特殊设计的闸波波形与低通滤波器,并将其用于副载波锁相环的相关过程中,这一思路与码环的闸波设计具有相似之处。通过采用闸波参与相干积分,本地信号与接收信号的相关函数发生了改变,其非零区域减小,进而提高了对副载波多径误差的抑制能力。

从信号跟踪的两个方面对该算法进行了评估,一方面是考虑其多径抑制能力,另一方面则是考虑其热噪声性能。通过仿真分析发现改进的双载波环路法相比于传统的双载波环路法会带来 -6 dB 的相干后载噪比损失,其原因是算法通过改变相关过程,本质上缩短了相干积分长度从而降低了能量积累。但是改进的算法相比于双载波环路法,可以显著地改善副载波多径误差包络,对于 BOC(1, 1)信号,其降低了 81.1% 的副载波多径误差包络面积;对于 BOC(14, 2)信号,其降低了 75.1% 的副载波多径误差包络面积。因此综合来看,本算法对闸波波形的设计需要权衡考虑,以平衡算法的多径抑制能力和噪声性能。相比于其他接收算法,本算法更适用于非弱信号条件下的高阶 BOC 信号抗多径接收。

参考文献(References)

- [1] Lee Y C. Compatibility of the new military GPS signals with non-aviation receivers [C]//Proceedings of the 58th Annual Meeting of the Institute of Navigation ION-AM, Albuquerque, NM, 2002: 581 - 597.
- [2] Fine P, Wilson W. Tracking algorithm for GPS offset carrier signals [C]//Proceedings of the US Institute of Navigation 1999 National Technical Meeting (NTM) Conference, San Diego, CA, 2007: 1017 - 1027.
- [3] Fishman P M, Betz J W. Predicting performance of direct acquisition for the M-code signal [C]//Proceedings of the US Institute of Navigation 2000 National Technical Meeting (NTM) Conference, Anaheim, CA, 2000: 574 - 582.
- [4] Dovis F, Mulassano P, Presti L L. A novel algorithm for the code tracking of BOC(n, n) modulated signals [C]//Proceedings of the 18th the International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS), Long Beach, CA, 2005: 152 - 157.
- [5] Hodgart M S, Blunt P, Unwin M. The optimal dual estimate solution for robust tracking of binary offset carrier (BOC) modulation [C]//Proceedings of the 20th International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS), Fort Worth, TX, 2007: 1017 - 1027.
- [6] Borio D. Double phase estimator: new unambiguous binary offset carrier tracking algorithm [J]. IET Radar, Sonar and Navigation, 2014, 8(7): 729 - 741.
- [7] Borio D. Double phase estimator: new results [C]//Proceedings of the Satellite Navigation Technologies and European Workshop on GNSS Signals and Signal Processing (NAVITEC), 7th ESA Workshop on IEEE, Noordwijk, Netherlands, 2014: 1 - 6.
- [8] Garin L J. The "shaping correlator", novel multipath mitigation technique applicable to GALILEO BOC(1, 1) modulation waveforms in high volume markets [C]//Proceedings of the ENC-GNSS 2005, Munich, Germany, 2005: 1 - 16.
- [9] Axelrad P, Larson K, Jones B. Use of the correct satellite repeat period to characterize and reduce site-specific multipath errors [C]//Proceedings of the 18th the International Technical Meeting of the Satellite Division of the Institute of Navigation (ION GNSS), Long Beach, CA, 2005: 2638 - 2648.
- [10] Maqsood M, Gao S, Brown T, et al. A compact multipath mitigating ground plane for multiband GNSS antennas [J]. IEEE Transactions on Antennas and Propagation, 2013, 61(5): 2775 - 2781.
- [11] Btaille D, Maenpa J, Euler E, et al. A new approach to GPS phase multipath mitigation [C]//Proceedings of the US Institute of Navigation NTM Conference, Anaheim, CA, 2013: 152 - 157.

应用于卫星导航功率倒置阵的改进最小均方算法*

陈飞强, 聂俊伟, 倪少杰, 王飞雪
(国防科技大学 电子科学与工程学院, 湖南 长沙 410073)

摘要:采用功率倒置准则的自适应天线阵特别适合于弱信号、强干扰的场合,因而在卫星导航系统中得到了广泛的应用。针对基于最小均方算法实现的卫星导航功率倒置阵在干扰数目或干扰功率突然减少时,算法收敛慢、影响信号接收性能的问题,分析了这一现象的产生机理,并提出了相应的改进算法。改进算法通过功率监测来检测干扰数目或干扰功率的突变,然后对最小均方算法进行复位处理重置权值来达到迅速收敛的目的。仿真结果表明:与原算法相比,改进算法可显著提高功率倒置阵的收敛速度。

关键词:卫星导航;功率倒置;天线阵;抗干扰;最小均方

中图分类号:TN967.1 **文献标志码:**A **文章编号:**1001-2486(2017)03-047-05

Improved least mean square algorithm for power-inversion global navigation satellite system antenna array

CHEN Feiqiang, NIE Junwei, NI Shaojie, WANG Feixue
(College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: The power-inversion adaptive array is very suitable for situations where the desired signals are very weak while the interfering signals are much stronger, thus, it is widely used in the global navigation satellite system. For least mean square based on power-inversion adaptive array which is used in global navigation satellite system receivers, the convergence rate was very slow when sudden change in interference number or interference power occurs. The performance of receiving the satellite signals would be degraded. The cause of this phenomenon was analyzed. Then, an improved least mean square algorithm was proposed to solve this problem. The key novelty of the proposed method was that it monitors the power of the reference antenna output to detect any sudden change in interference number or interference power. Once the interference number or interference power decreases suddenly, the weight vector would be initialized again to improve the rate of convergence. Simulation results show that the improved method outperforms the original algorithm in rate of convergence.

Key words: satellite navigation; power-inversion; antenna array; anti-jamming; least mean square

对于全球导航卫星系统(Global Navigation Satellite System, GNSS),由于到达地面的卫星信号十分微弱(通常比热噪声小于30 dB),GNSS接收机极易被干扰。干扰会导致接收端的信噪比下降,从而使定位授时精度恶化,甚至使得接收机完全无法工作^[1]。

自适应天线阵是一种有效的GNSS抗干扰措施^[2-6],它通过控制阵列中各阵元的增益和相位,使阵列方向图在干扰方向形成零陷来抑制干扰。经典的阵列加权准则包括最小方差无失真响应(Minimum Variance Distortionless Response, MVDR)准则^[7-8]、最小均方误差(Minimum Mean Square Error, MMSE)准则^[9]和功率倒置(Power Inversion, PI)准则^[10-11]等。其中PI准则是一种盲抗干扰准则,不需要先验信息辅助,因而可以低成本在一个独立的抗干扰硬件单元中实现;通用GNSS接收机不需要做任何修改即可与其直接相连完成抗干扰接收功能^[12]。这些特点使得功率倒置阵在GNSS抗干扰接收机中得到了广泛的应用。

PI准则的基本原理是以某一个阵元接收信号作为参考,调整其他支路的阵列取值使阵列的误差输出信号功率最小。阵列权值可通过最小均方(Least Mean Square, LMS)算法获得,LMS算法具有算法复杂度小、易于工程实现等优点。但是,在干扰数目或干扰功率突然减少的情况下,其收敛速度非常慢,以至于这些干扰对应的零陷不能迅速消失或变浅。若卫星信号入射方向接近这些零陷方向,其结果必然会导致卫星信号出现较大

* 收稿日期:2015-12-30
基金项目:国家自然科学基金资助项目(61371158,61071140)
作者简介:陈飞强(1988—),男,湖南益阳人,博士研究生,E-mail:matlabfly@hotmail.com;
王飞雪(通信作者),男,教授,博士,博士生导师,E-mail:wangfeixue365@sina.com

的功率损耗。本文对这一现象产生的机理进行了深入分析,并提出一种改进 LMS 算法来提高算法的收敛速度。

1 问题描述及机理分析

考虑一个 N 元天线阵,设 $x_r(n)$ 表示参考阵元接收的信号,这个信号一般也用作 LMS 算法的参考信号。其他 $N-1$ 个阵元接收的信号用 $x_k(n) (k = 1, 2, \dots, N-1)$ 表示,这些信号经空域滤波后的输出可表示为:

$$y_a(n) = \sum_{k=1}^{N-1} w_k^*(n) x_k(n) = \mathbf{w}_a^H(n) \mathbf{x}_a(n) \quad (1)$$

式中, $\mathbf{w}_a = [w_1(n), w_2(n), \dots, w_{N-1}(n)]^T$ 为阵列权值, $\mathbf{x}_a(n) = [x_1(n), x_2(n), \dots, x_{N-1}(n)]^T$, $(\cdot)^T$ 表示转置, $(\cdot)^H$ 表示共轭转置。

根据 PI 准则的原理,总的阵列输出,即参考信号与空域滤波输出信号之差,可表示为:

$$y(n) = e(n) = x_r(n) - y_a(n) \quad (2)$$

采用 LMS 算法,通过使阵列输出的均方值最小可得到阵列权值的更新表达式:

$$\begin{aligned} \mathbf{w}_a(n+1) &= \mathbf{w}_a(n) - \mu \mathbf{x}_a(n) e^*(n) \\ &= \mathbf{w}_a(n) + \Delta \mathbf{w}_a(n) \end{aligned} \quad (3)$$

式中, μ 为步长因子, $\Delta \mathbf{w}_a(n)$ 为权值增量。

不失一般性,考虑有 M 个干扰的场景 ($M < N$),当 LMS 算法收敛时,天线阵将在这 M 个干扰的入射方向形成相应的 M 个零陷来抑制干扰。同时,误差信号 $e(n)$ 的功率趋近最小值。由于干扰被抑制,因此误差信号主要由热噪声组成。由于算法处于收敛状态,此时权值增量的幅度相对权值本身的幅度来说非常小,即有:

$$\mathbf{w}_a(n) + \Delta \mathbf{w}(n) \approx \mathbf{w}_a(n) \quad (4)$$

当干扰数目由 M 个增加到 $M+K$ ($M+K < N$) 个时,则此时的阵列权值无法抑制新出现的 K 个干扰。误差信号的功率将会迅速增大,从式(3)可知,权值增量的幅度也将相应增大,从而使阵列权值迅速收敛到新的稳定状态(即天线阵形成 $M+K$ 个零陷)来抑制新出现的干扰。

然而,当干扰数目减少 K 个时,由于剩下的 $M-K$ 个干扰仍然被天线阵形成的零陷所抑制,此时误差信号仍然由热噪声组成,功率不会发生明显的改变。因此,在这种条件下式(4)仍然成立,权值增量的幅度相对权值本身的幅度来说非常小,算法需要很长的时间才能收敛到新的稳定状态(即天线阵形成 $M-K$ 个零陷)。其结果是

这 K 个干扰对应的零陷不能迅速消失,若卫星信号入射方向接近这些零陷方向,其结果必然会导致卫星信号出现较大的功率损耗。

干扰功率突然变小的情况与干扰数目减少的情况类似,由于算法收敛很慢,其结果是干扰对应的零陷不能迅速变浅,也将影响到卫星信号的正常接收。

根据上面的分析可知,在干扰数目或干扰功率突然减少时,误差信号的功率并不会发生明显改变,误差信号功率过小是导致算法收敛速度慢的主要原因。

2 算法改进

针对上述问题,提出一种改进的 LMS 算法,该算法的核心思想是通过参考阵元接收的信号进行功率监测来检测干扰数目或干扰功率的突变。一旦检测到干扰数目或干扰功率突然减少,则对 LMS 算法的迭代过程进行复位处理,从而达到使算法迅速收敛的目的。

参考阵元接收信号的功率可通过式(5)进行估计。

$$p(m) = \frac{1}{L} \sum_{n=mL+1}^{mL+L} x_r^*(n) x_r(n) \quad (5)$$

式中, L 为用于估计功率的数据块长度。

所提算法的实现框图如图 1 所示。在权值迭代之前,首先要对阵列权值进行初始化处理,即令

$$\mathbf{w}_a(1) = \mathbf{w}_0 \quad (6)$$

其中, \mathbf{w}_0 为一个任意的 $N-1$ 维列向量,一般可取零向量作为初始值。

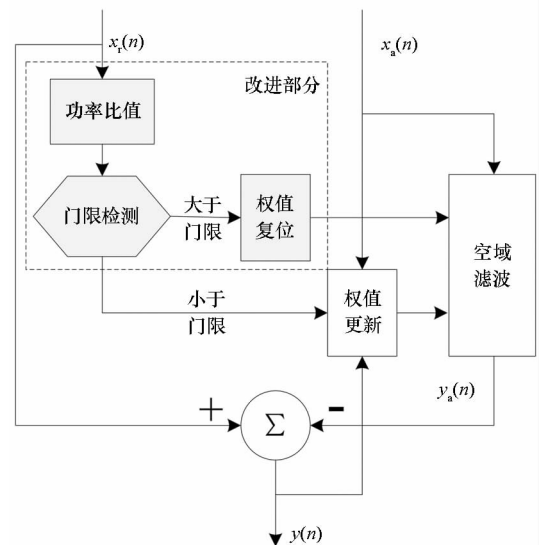


图 1 算法实现框图

Fig. 1 Block diagram of proposed algorithm

用阵列权值 $\mathbf{w}_a(n)$ 对阵列输入数据 $\mathbf{x}_a(n)$ 进

行空域滤波处理(见式(1))可得到空域滤波输出 $y_a(n)$ 。空域滤波输出与参考信号 $x_r(n)$ 作差即可得到最终的阵列输出信号 $y(n)$ 。

对于阵列权值的计算,经典的LMS算法直接根据误差信号(在这里为阵列输出信号)和阵列输入信号即可计算出权值增量,并进一步可用式(3)进行权值更新,得到下一个快拍对应的阵列权值。

在经典LMS算法的基础上进行了相应的改进,改进部分如图1中的虚线框所示。首先用式(5)对参考阵元接收信号的功率进行块估计,得到前一个数据块对应的功率 $p(m-1)$ 与当前数据块对应的功率 $p(m)$ 的比值,并进行门限检测,即判断式(7)是否成立。

$$\frac{p(m-1)}{p(m)} > T$$

(7)

式中, T 为检测门限。门限检测的目的是检测干扰数目或干扰功率的突然减少。为了降低虚警概率,可采取提高门限值或使用双门限检测等措施,这一部分的内容可参考文献[13]。

当检测量小于门限值时,根据式(3)对阵列权值进行更新。而当检测量大于门限值时,对阵列权值进行复位处理,即令

$$\mathbf{w}_a(n+1) = \mathbf{w}_0$$

(8)

复位处理后,误差信号会突然增大,相应的权值增量也迅速变大,从而使得阵列权值迅速收敛到新的稳定状态。

相对传统的LMS算法,改进算法增加了一个功率监测模块,因此会增加算法实现复杂度。功率监测模块增加了一次复数乘法和一次复数加法运算,但这些运算量相对整个抗干扰算法来说非常小,不会影响到算法的实时性。

3 性能仿真

为了验证算法的性能,用软件接收机进行仿真。首先用MATLAB生成阵列信号,用来模拟产生天线阵接收到的不同入射方向上的GNSS信号、干扰以及噪声。然后用该算法对生成的阵列信号进行处理,并与经典的LMS算法进行对比。基本的仿真参数设置见表1。

仿真实验一中,在 $t=0$ 时刻,同时开启两个干扰,在 $t=50\text{ ms}$ 时关闭干扰1,用于模拟干扰数目的突然减少。仿真中,数据块长度 $L=1024$,检测门限 T 取经验值 $1.28^{[14]}$ 。

表1 仿真中用到的参数	
Tab. 1 Parameters used in simulations	
参数类型	参数取值
天线阵型	间距为半波长的四元直线阵
GNSS 信号类型	北斗 B3 一期民码信号 (PRN 1)
信号入射角	30° (线阵法线方向对应 0°)
信噪比	-30 dB (热噪声功率设为 0 dB)
干扰 1 类型	B3 频点的单频干扰
干扰 1 干信比	60 dB
干扰 1 入射角	20°
干扰 2 类型	20 MHz 宽带高斯干扰
干扰 2 干信比	65 dB
干扰 2 入射角	45°

图2给出了不同时刻下LMS算法得到的天线阵方向图,其中算法的步长因子固定为 10^{-6} 。从图2中可以看到,天线阵在两个干扰方向均形成了零陷来抑制干扰,由于卫星信号的入射方向与干扰1的零陷方向接近,因此卫星信号也被部分抑制。在第50 ms关闭干扰1后,天线阵在20°方向处的零陷并没有迅速消失,而是慢慢变浅,直到第100 ms时才基本消失。干扰数目突然减少后,整个算法的收敛过程十分缓慢。这个实验结果与第1节的理论分析一致。

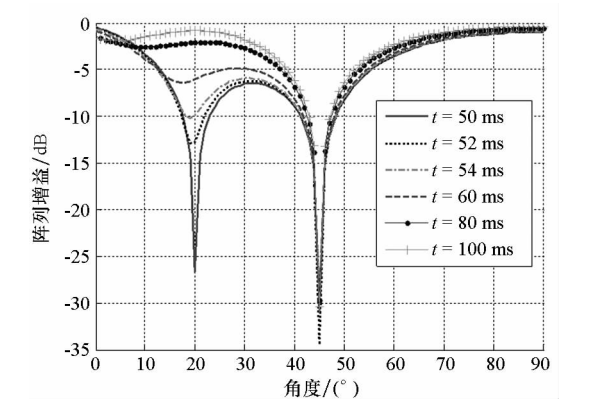


图2 LMS算法得到的天线阵方向图(实验一)
Fig. 2 Antenna pattern obtained by LMS algorithm (Scenario 1)

图3给出了不同时刻下算法得到的天线阵方向图。从图3中可以看到,当干扰1在第50 ms关闭后,天线阵在20°方向处的零陷迅速消失,在第51 ms时算法就基本收敛,这时卫星信号的功率损耗也可以立即恢复。

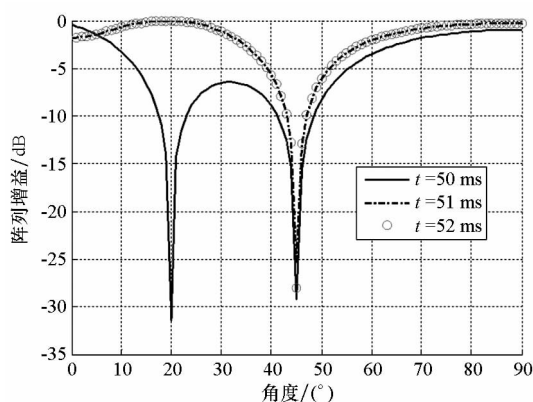


图 3 本文算法得到的天线阵方向图(实验一)

Fig. 3 Antenna pattern obtained by proposed algorithm (Scenario 1)

图 4 进一步给出了不同算法得到的阵列输出信干噪比 (Signal to Interference plus Noise Ratio, SINR), 其定义为阵列输出信号功率与干扰加噪声功率之比。从图 4 中可以看到, 当干扰 1 关闭时, 本文算法的收敛速度明显优于 LMS 算法和归一化最小均方 (Normalized LMS, NLMS) 算法^[15]。实际上, 本文算法在干扰 1 关闭后迅速收敛到了新的稳定状态 (即天线阵只在 45° 方向形成零陷), 从而阵列输出信干噪比迅速提升。实验中 NLMS 算法的步长因子固定为 0.05。值得注意的是, 对于 LMS 算法和 NLMS 算法, 增大步长因子可以提高算法的收敛速度, 但也会同时增大失调, 从而使算法性能下降。

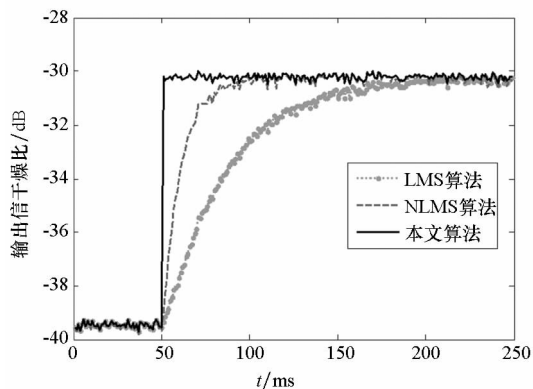


图 4 不同时刻下的阵列输出信干噪比(实验一)

Fig. 4 Output SINR against times (Scenario 1)

仿真实验二中, 在 $t=0$ 时刻, 开启两个干扰, 在 $t=50$ ms 时将两个干扰的功率衰减 20 dB, 用于模拟干扰功率由于遮挡等原因引起的突然衰减。仿真实验中其他参数与仿真实验一的相同。

图 5 和图 6 对比了不同时刻下 LMS 算法和本文算法得到的天线阵方向图。从图 5 中可以看出, 对于 LMS 算法, 当干扰功率在第 50 ms 衰减

时, 天线阵在干扰方向的零陷并没有迅速变浅, 而是慢慢变浅, 整个算法的收敛过程十分缓慢。从图 6 中可以看出, 对于本文算法, 天线阵在干扰方向的零陷迅速变浅, 在第 51 ms 时算法就基本收敛, 实验结果与理论分析一致。

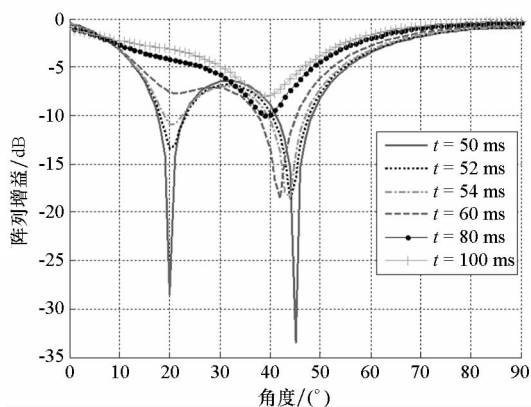


图 5 LMS 算法得到的天线阵方向图(实验二)

Fig. 5 Antenna pattern obtained by LMS algorithm (Scenario 2)

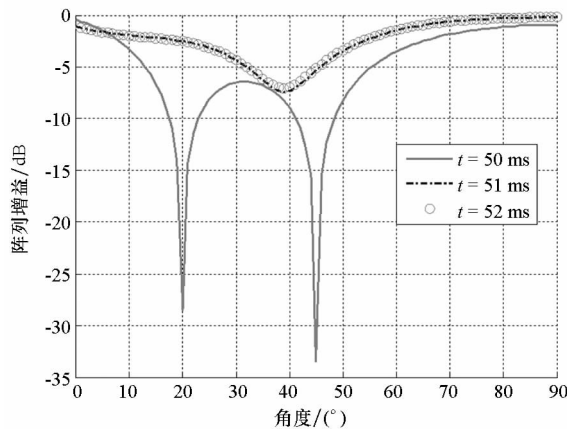


图 6 本文算法得到的天线阵方向图(实验二)

Fig. 6 Antenna pattern obtained by proposed algorithm (Scenario 2)

4 结论

针对基于 LMS 算法实现的卫星导航功率倒置阵, 在干扰数目或干扰功率突然减少时算法收敛慢、影响信号接收的问题, 首先从理论上对这一现象的机理进行了分析, 指出误差信号功率过小是导致算法收敛慢的主要原因。进一步, 提出了一种改进的 LMS 算法, 该算法通过功率监测来检测干扰数目和干扰功率的突变, 并通过复位处理重置权值来达到使算法迅速收敛的目的。仿真实验结果表明, 与 LMS 算法和 NLMS 算法相比, 该方法可显著提高算法的收敛速度。该研究成果对卫星导航功率倒置阵的工程实现有一定的指导价值。

参考文献 (References)

[1]

Kaplan E D, Hegarty C J. Understanding GPS: principles and applications [M]. 2nd ed. US: Artech House, 2006.

[2]

O'Brien A J. Adaptive antenna arrays for precision GNSS receivers [D]. US: The Ohio State University, 2009.

[3]

Arribas J, Fernandez-Prades C, Closas P. Antenna array based GNSS signal acquisition for interference mitigation [J]. IEEE Transactions on Aerospace and Electronic Systems, 2013, 49(1): 223 – 243.

[4]

Li M, Dempster A G, Balaei A T, et al. Switchable beam steering/null steering algorithm for CW interference mitigation in GPS C/A code receivers [J]. IEEE Transactions on Aerospace and Electronics Systems, 2011, 47(3): 1564 – 1579.

[5]

Daneshmand S, Broumandan A, Lachapelle G. GNSS interference and multipath suppression using an antenna array[C]//Proceedings of the 24th International Technical Meeting of the Institute of Navigation, 2011: 1183 – 1192.

[6]

Seco-Granados G, Fernandez-Rubio J A, Fernandez-Prades C. ML estimator and hybrid beamformer for multipath and interference mitigation in GNSS receivers [J]. IEEE Transactions on Signal Processing, 2005, 53(3): 1194 – 1208.

[7]

Applebaum S P. Adaptive arrays [J]. IEEE Transaction on Antennas and Propagation, 1976, 24(5): 585 – 598.

[8]

Zhang Y D, Amin M G. Anti-jamming GPS receiver with reduced phase distortions [J]. IEEE Signal Processing Letters, 2012, 19(10): 635 – 638.

[9]

Widrow B. Adaptive antenna systems [J]. Proceedings of the IEEE, 1967, 55(12): 2143 – 2159.

[10]

桑怀胜, 李峥嵘, 王飞雪, 等. 采用 RLS 算法的功率倒置阵列的性能 [J]. 国防科技大学学报, 2003, 25(3): 36 – 40.
SANG Huaisheng, LI Zhengrong, WANG Feixue, et al. The performance of power inversion array using RLS algorithm[J]. Journal of National University of Defense Technology, 2003, 25(3): 36 – 40. (in Chinese)

[11]

Compton R T, Jr.. The power-inversion adaptive array: concept and performance [J]. IEEE Transactions on Aerospace and Electronic Systems, 1979, 15(6): 803 – 813.

[12]

Fu Z, Hornbostel A, Hammesfahr J, et al. Suppression of multipath and jamming signals by digital beamforming for GPS/Galileo applications[J]. GPS Solution, 2003, 6(4): 257 – 264.

[13]

Kay S M. Fundamentals of statistical signal processing [M]. US: Prentice Hall PTR, 1998.

[14]

王瑛. 卫星导航天线阵抗干扰关键技术研究[D]. 长沙: 国防科技大学, 2008.
WANG Ying. Research on the key technologies of anti-jamming antenna arrays in satellite navigation systems[D]. Changsha: National University of Defense Technology, 2008. (in Chinese)

[15]

Goodwin G C, Sin K S. Adaptive filtering, prediction, and control [M]. US: Prentice-Hall, 1984.

[3]

Cai C S, Gao Y. Performance analysis of precise point positioning based on combination GPS and GLONASS [C]//Proceedings of the ION GNSS 20th International Technical Meeting of the Satellite Division, Fort Worth, Texas, 2007.

[4]

黄令勇, 吕志平, 任雅奇, 等. 多元总体最小二乘在三维坐标转换中的应用[J]. 武汉大学学报: 信息科学版, 2014, 39(7): 793 – 798.
HUANG Lingyong, LYU Zhiping, REN Yaqi, et al. Application of multivariate total least square in three-dimensional coordinate transformation [J]. Geomatics and Information Science of Wuhan University, 2014, 39(7): 793 – 798. (in Chinese)

[5]

黄令勇, 翟国君, 欧阳永忠, 等. 三频 GNSS 电离层周跳处理[J]. 测绘学报, 2015, 44(7): 717 – 725.
HUANG Lingyong, ZHAI Guojun, OUYANG Yongzhong, et al. Ionospheric cycle slip processing in triple-frequency GNSS[J]. Acta Geodaetica et Cartographica Sina, 2015, 44(7): 717 – 725. (in Chinese)

[6]

黄令勇, 翟国君, 欧阳永忠, 等. 削弱电离层延迟影响的三频 TurboEdit 周跳处理方法[J]. 测绘学报, 2015, 44(8): 840 – 847.
HUANG Lingyong, ZHAI Guojun, OUYANG Yongzhong, et al. Triple-frequency TurboEdit cycles lip processing method of weakening ionospheric activity [J]. Acta Geodaetica et Cartographica Sina, 2015, 44(8): 840 – 847. (in Chinese)

[7]

杨元喜. 自适应动态导航定位[M]. 北京: 测绘出版社, 2006.
YANG Yuanxi. Adaptive navigation and kinematic positioning[M]. Beijing: Surveying and Mapping Press, 2006. (in Chinese)

(上接第 35 页)

水下应用栅格翼动态展开参数预示方法*

鲍文春, 权晓波, 李 岩, 程少华, 王占莹
(北京宇航系统工程研究所, 北京 100076)

摘 要: 基于水下应用栅格翼动态展开过程动力学模型, 根据展开特征角度定常水动力方法, 拟合获得展开全程受到的流体力矩, 并引入考虑相对速度影响的修正因子, 形成水下应用栅格翼动态展开过程参数工程预示方法, 以典型展开时序点航行体运动参数为设计输入, 对栅格翼展开过程运动参数进行预示。通过与水下应用栅格翼非定常流场仿真计算数据以及水下航行体弹射试验数据对比, 验证了上述预示方法的正确性及工程适用性, 为水下应用栅格翼方案设计的优化及展开不同步性分析提供设计参考。

关键词: 栅格翼; 动态展开; 预示方法

中图分类号: V211.3 **文献标志码:** A **文章编号:** 1001-2486(2017)03-052-06

Parameters prediction method of underwater grid fins during the procedure of dynamic expansion

BAO Wenchun, QUAN Xiaobo, LI Yan, CHENG Shaohua, WANG Zhanying
(Beijing Institute of Space System Engineering, Beijing 100076, China)

Abstract: Based on the kinetic model of underwater grid fins during the procedure of dynamic expansion, a prediction method was developed, which took the motion parameters at typical expansion points as the design input. The hydrodynamic moment used in the kinetic model was fitted by the values obtained in steady hydrodynamic conditions with specific expansion angles. A modified factor in consideration of the relative velocity was included as well. By comparing with the results obtained by unsteady numerical simulations and the underwater vehicle experiment data, the prediction method is verified, which can provide a design reference for the optimum proposal of the underwater grid fins and the analysis of non-synchronization expansion.

Key words: grid fin; dynamic expansion; prediction method

栅格翼是由很多薄的翼片镶嵌在边框内组成的升力体系统。近年来, 栅格翼作为一种新型承力结构凭借其尺寸较小、可折叠安装、升力面积大、强度-重量比高等优点被广泛应用于航天及武器型号研制之中^[1]。借鉴栅格翼在改善气动力方面的优势, 拟将栅格翼技术应用于水下航行体水下发射技术中。受发射方案及发射筒空间的限制, 栅格翼在发射筒内须处于折叠状态^[2], 航行体出筒后, 栅格翼在展开机构及本身的水动力矩作用下自动打开, 在航行体水中运动阶段起到稳定和控制在作用^[3]。由于水的密度约为空气密度的 800 倍, 栅格翼展开过程中角速度不断增加, 因此栅格翼展开到位后将引起较强的冲击, 对展开角速度进行预示是栅格翼动态强度设计的先决条件。水下应用栅格翼在展开结构作用下完成初始展开动作, 然后在自身流体动力作

用下持续展开动作。由于水下垂直发射航行体水中作用时间较短, 因此水下应用栅格翼展开用时是水下栅格翼方案的关键设计指标之一。完成特定展开方案下栅格翼展开过程参数预示, 将为分析栅格翼动态强度设计、栅格翼展开机构设计, 以及展开时序等的合理性及可行性奠定基础, 同时也是航行体上多片栅格翼展开不同步性分析的重要依据。

目前, 针对水下应用栅格翼动态展开问题相关的可参考文献较少, 理论研究也需要进一步完善。

1 栅格翼动态展开动力学模型

本节首先选取一定外形尺寸的栅格翼, 建立其展开过程动力学模型。利用定常计算结果得到展开过程期间栅格翼受到的流体力矩, 求解动态

* 收稿日期: 2015-12-25
基金项目: 教育部重点实验室基金资助项目(11172325, 90716015)
作者简介: 鲍文春(1988—), 女, 吉林榆树人, 工程师, 硕士, E-mail: baowenchun@126.com

展开过程控制方程组,获得栅格翼动态展开角速度等参数变化规律。形成适用于栅格翼动态展开过程参数工程预示方法。

1.1 栅格翼外形参数

根据已有文献中栅格翼结构特征^[4],选择如图1所示的斜置蜂窝状栅格翼进行研究,栅格翼翼形剖面为矩形,具体尺寸参数如表1所示。

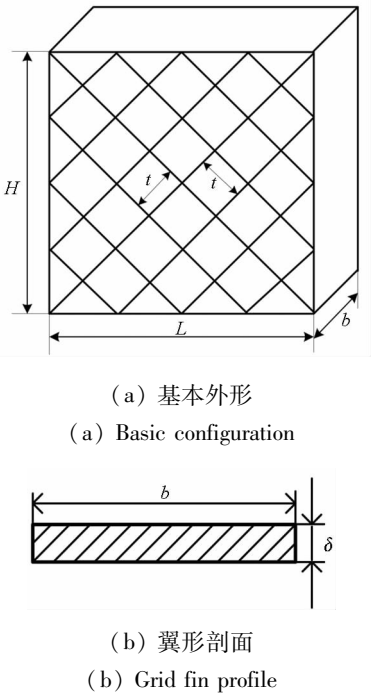


图1 栅格翼几何外形示意图
Fig.1 Schematic of the configuration of the grid fin

表1 栅格翼结构尺寸

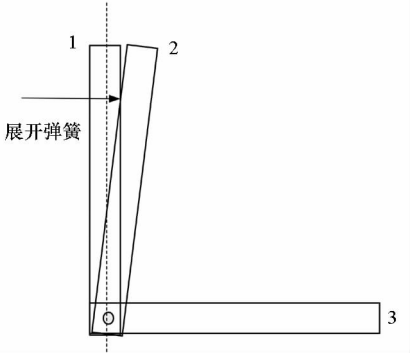
Tab.1 Structure size of the grid fin mm				
翼高 H	翼展 L	翼弦 b	格间距 t	翼片厚度 δ
210	220	25	30	2

1.2 栅格翼展开过程控制方程组

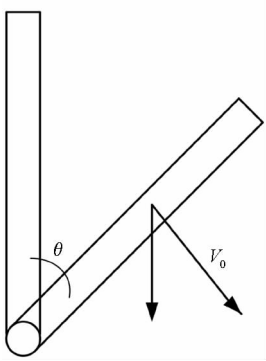
图2给出了水下应用栅格翼展开过程示意图,其中 V_0 代表栅格翼展开时航行体运动速度。栅格翼在筒内折叠安装在模型尾部区域(位置1),出筒后栅格翼解锁在助展弹簧和水动外力的作用下展开至与弹体垂直的设计位置(位置3)。根据栅格翼展开过程受力特性将展开过程分为两个阶段进行分析。

阶段1:栅格翼解锁至边框上沿突出弹体表面(位置1~位置2,此时 $\theta \leq 5^\circ$),此过程栅格翼运动速度较低,翼片前后面压差较小,栅格翼受到的流体力矩较小,此过程栅格翼主要在展开弹簧、转轴摩擦力等的作用下开始绕转轴运动;

阶段2:栅格翼边框突出弹体至展开到位过程(位置2~位置3,此时 $\theta \leq 5^\circ$),此过程栅格翼在水动外力矩的作用下加速展开。



(a) 栅格翼在展开弹簧作用下初始展开
(a) Initial expansion by the spring mechanism



(b) 栅格翼在水动力作用下继续展开
(b) Further expansion by the hydrodynamic force

图2 栅格翼展开过程示意图

Fig.2 Schematic of the expansion process

根据栅格翼展开过程受力分析并参考相关文献资料^[2,5],可以建立栅格翼绕转轴展开过程运动方程:

$$\frac{d\omega}{dt} = \begin{cases} (M_{\text{簧}} - M_{\text{摩擦阻力矩}})/(J_z + \lambda_{66}), & \text{位置1~位置2} \\ (M_{\text{水外力矩}} - M_{\text{摩擦阻力矩}})/(J_z + \lambda_{66}), & \text{位置2~位置3} \end{cases} \quad (1)$$

其中, ω 为栅格翼展开角速度, $M_{\text{簧}}$ 为展开弹簧驱动力矩,与弹簧刚度 K ,助展弹簧作用力臂 l 和弹簧压缩量 Δx 相关:

$$M_{\text{簧}} = K \times \Delta x \times l \quad (2)$$

根据试验设计的展开机构,实际应用中 $M_{\text{簧}}$ 可以通过以下方程计算:

$$M_{\text{簧}} = 17 \times (18 - 182\theta) \times 0.182 = 55.69 - 563.1\theta \quad (3)$$

其中, θ 为栅格翼展开角度, θ 在 $0^\circ \sim 5^\circ$ 范围内 $M_{\text{簧}}$ 存在, θ 大于 5° 时展开弹簧与栅格翼不发生作用, $M_{\text{簧}} = 0$ 。 $M_{\text{摩擦阻力矩}}$ 为栅格翼转轴摩擦力矩,根据出筒时弹翼运动速度、转轴半径和相应材料

摩擦系数,运动过程中的摩擦力矩可表示为:

$$M_{\text{摩擦力矩}} = \frac{1}{2} \rho V^2 S \times f \times R \quad (4)$$

其中, V 为弹翼组合体运动速度, S 为参考面积, f 为摩擦因子, R 为转轴半径, 均可以根据设计方案直接获得。 $M_{\text{水外力矩}}$ 为栅格翼所受的水动外力矩。水动展开力矩可以通过定常计算方法得到不同展开角度下的力矩系数; 实际展开过程中, 由于栅格翼的转动, 栅格翼实际相对水流速度将减小, 见图 2。考虑栅格翼转动引起的运动速度影响, 展开过程分析中需要对定常状态计算得到的外力矩进行修正。定常状态展开下的力矩可以用 M_{z0} 表示, 考虑展开过程中的相对速度 $V' = V_0 - \omega R \sin \theta$, 引入力矩修正系数如式 (5) 所示。

$$K_{\theta} = \left(\frac{V_0 - \omega R \sin \theta}{V_0} \right)^2 \quad (5)$$

因此, 展开过程分析中水动外力矩通过式 (6) 计算:

$$M_{\text{水外力矩}} = K_{\theta} M_{z0} \quad (6)$$

式 (1) 中的 J_z 为栅格翼绕转轴的转动惯量, 可通过栅格翼具体结构计算获得; λ_{66} 为栅格翼附加转动惯量, 可由基于雷诺平均纳维-斯托克斯 (Reynolds Average Navier-Stokes, RANS) 方程的全黏流的附加质量计算^[6] 获得。

综上所述, 可以得到栅格翼展开过程动力学控制模型为:

$$\begin{cases} \frac{d\omega}{dt} = \left\{ \begin{array}{l} (M_{\text{簧}} - M_{\text{摩擦力矩}}) / (J_z + \lambda_{66}) \\ (M_{\text{水外力矩}} - M_{\text{摩擦力矩}}) / (J_z + \lambda_{66}) \end{array} \right. \\ M_{\text{簧}} = K \times \Delta x \times l \\ M_{\text{摩擦力矩}} = \frac{1}{2} \rho V^2 S \times f \times R \\ M_{\text{水外力矩}} = K_{\theta} M_{z0} \\ K_{\theta} = \left(\frac{V_0 - \omega R \sin \theta}{V_0} \right)^2 \end{cases} \quad (7)$$

2 基于定常力系数求解栅格翼展开控制方程组

栅格展开过程动力学控制方程组中仅展开过程中的水动外力矩为未知量, 为此建立栅格翼展开特征角度定常力计算数值仿真模型, 获得栅格翼不同展开特征角度位置时刻受到的流体力矩, 进而拟合获得展开全过程的流体力矩变化规律。选取栅格翼相对弹轴展开 30° 、 45° 、 60° 及 90° 特征阶段为研究对象, 建立相应的定常力计算模型并计算获得展开特征角度时的定常流体力矩。建

模过程中, 需要考虑弹体壁面对流场结构及栅格翼受力的影响。

在对航行体及栅格翼组合体进行网格划分时, 为了提高计算的精度, 保证栅格翼及航行体壁面附近的网格精度, 对整个计算域采用分区结构化网格的形式^[7], 栅格翼及航行体壁面附近的网格尺寸约为 $0.5 \sim 1 \text{ mm}$, 总网格数约为 800 万, 图 3 给出了栅格翼展开 60° 时展开流体力矩数值计算模型网格划分情况。

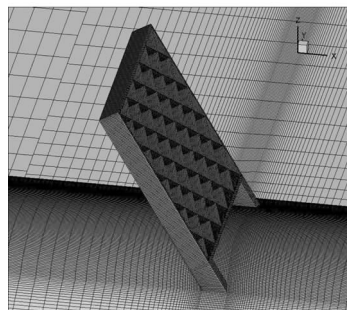


图 3 栅格翼展开 60° 角计算域内网格划分情况

Fig. 3 Calculation mesh as the expansion angle is 60°

根据相关参考文献^[8-10], 依据航行体栅格翼设计展开时序下的水下外压条件及航行体运动速度条件, 可以计算获得栅格翼展开过程特征角度定常水动力矩。求解过程中, 取力矩参考点为栅格翼相对转轴的中心点。考虑到计算结果的通用性, 以展开力矩系数的形式对栅格翼展开过程中的流体力矩进行分析, 展开力矩系数定义为:

$$C_{M_z} = \frac{M_z}{\frac{1}{2} \rho V_{\infty}^2 S L} \quad (8)$$

其中, M_z 为栅格翼受到的流体力矩, S 为参考面积, L 为参考长度, ρ 为水的密度, V_{∞} 为栅格翼展开时航行体运动速度。

图 4 为计算获得的栅格翼展开不同角度时受到的流体力矩系数。可见栅格翼展开过程中受到的流体力矩随展开角度逐渐增加, 栅格翼展开 90° 时, 展开力矩系数最大, 约为 0.000 95。

图 4 表明栅格翼展开过程中受到的流体力矩与展开角度相关, 通过对展开过程的理论分析, 将展开力矩拟合成展开角度的函数, 且有如下拟合公式:

$$\begin{cases} C_{M_z} \theta = K \times \sin \theta \\ C_{M_{z\min}} = 0 \\ C_{M_{z\max}} = C_{M_z} 90^\circ \end{cases} \quad (9)$$

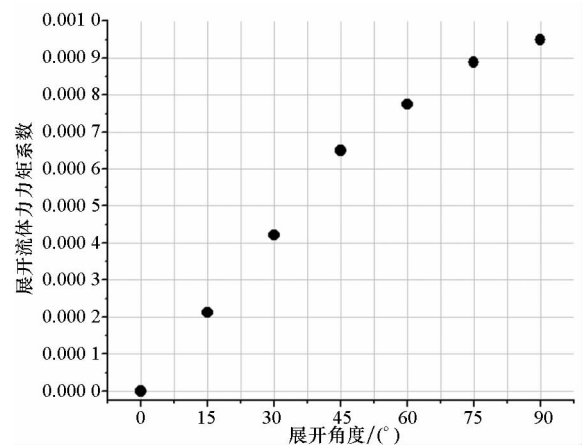


图4 展开力矩随展开角度变化规律

Fig.4 Variation of the expansion moment with expansion angles

代入初始边界条件及展开到位边界条件,可获得如下的拟合公式:

$$C_{M_z}\theta = \frac{M_z}{\frac{1}{2}\rho V_\infty^2 SL}$$
$$= C_{M_z}90^\circ \times \sin\theta = 4.25 \times 10^{-4} \times \sin\theta \quad (10)$$

图5中给出了拟合结果与特征角度计算结果的对比。从图中可以看出,应用拟合公式计算值与数值仿真得到的定常力系数符合较好,可以将拟合得到的力矩作为栅格翼动态展开过程中受到的流体作用力矩。

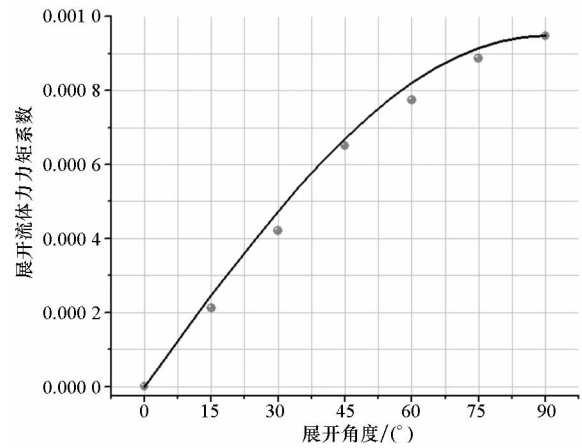


图5 拟合结果与计算值对比情况

Fig.5 Comparison of the fitting and numerical results

将展开流体力矩代入栅格翼展开过程动力学控制方程组(7),可以计算得到展开角度及角速度的变化规律,如图6、图7所示。根据计算结果,展开到位时刻的角速度为47.84 rad/s,展开到位所用时间为107.7 ms。

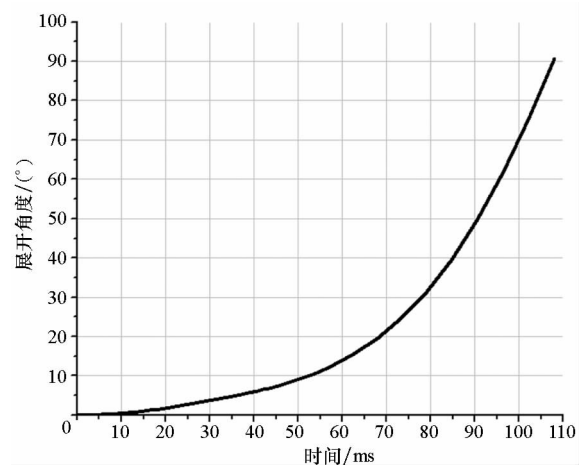


图6 栅格翼展开角度计算结果

Fig.6 Results of the calculated expansion angles

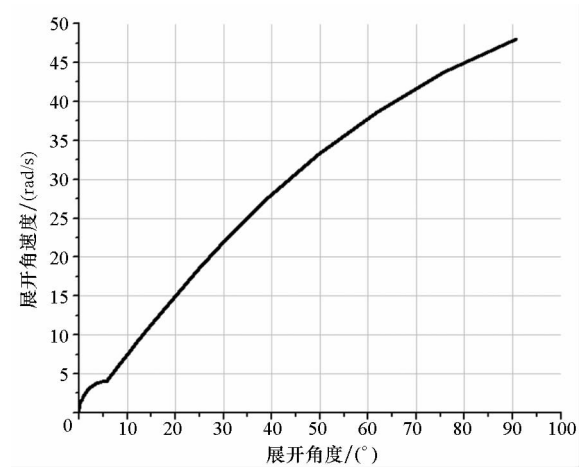


图7 栅格翼展开角速度计算结果

Fig.7 Results of the calculated expansion angular velocity

3 栅格翼动态展开参数工程预示方法验证

为了验证上述求解栅格翼展开过程动态参数预示方法,将计算结果分别与非定常数值模拟结果以及缩比模型航行体水下弹射试验结果进行对比,验证工程预示方法的正确性及适用性。

3.1 数值水洞非定常计算结果对比

依据水下栅格翼结构参数及展开机构设计结果,通过对商用数值计算软件Fluent的二次开发,可以对栅格翼水下展开过程进行数值模拟。在数值仿真求解时,采用流场与运动耦合的求解方式,即每个时间步内,先计算得到栅格翼受到的流体水动力矩,将力矩计算结果与展开机构的作用力矩叠加,获得栅格翼受到的展开作用力矩,进而求得栅格翼当前时刻的运动参数,并采用动网格技术实现对运动过程的描述及计算网格的更新。

图8中给出了应用工程拟合公式和采用数值仿真得到的展开角度随展开时间的变化情况,展

开角速度随展开角度变化情况见图 9。对比结果表明,应用本文方法获得的展开过程参数数值及变化规律与非定常数值试验获得的参数变化规律符合较好。验证了本文所建立的栅格翼展开过程参数预示方法的正确性。

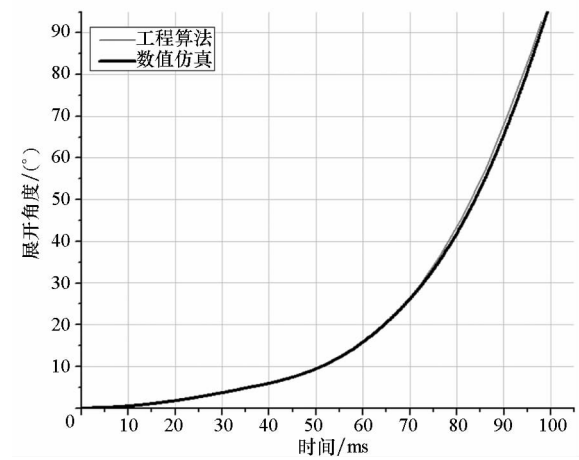


图 8 展开时间对比情况

Fig. 8 Comparison of the expansion time

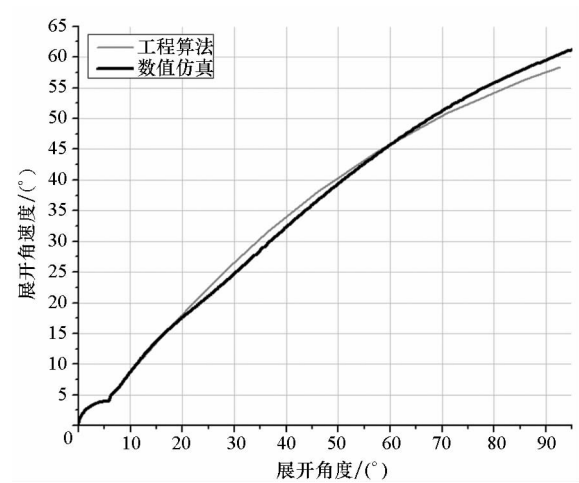


图 9 展开角速度对比情况

Fig. 9 Comparison of the expansion angular velocity

本文建立的栅格翼展开过程动力学模型以相对速度的方式对栅格翼展开期间非定常效应进行了修正;水下应用栅格翼在展开机构及流体力矩作用下展开,其中水动力矩起主要作用,而本文基于展开特征角度定常水动力矩计算结果,拟合获得了精度较高的栅格翼展开全程水动力矩数学表述,再将其代入展开过程动力学模型,因此求解获得的展开过程参数与非定常仿真求解结果吻合较好。

非定常数值仿真计算获得的栅格翼展开到位时刻的角速度为 53.64 rad/s,展开到位所用时间为 110.0 ms。以动态展开过程计算结果为准,两种算法得到的栅格翼展开到位时间相对误差为

-2.1%,展开到位时刻角速度相对误差为 -10.8%。

与栅格翼非定常展开数值模拟相比,本文所建立的动态展开参数预示方法省去了繁复耗时的水下流场与运动耦合数值仿真计算,可以快速获得展开过程参数变化规律及展开用时等特征点参数,为水下应用栅格翼展开结构设计及展开时序设计提供设计参考。

3.2 带栅格翼航行体水下弹射试验结果对比

通过弹翼组合体水下弹射试验,得到的单片栅格翼展开用时统计情况见表 2。弹射试验得到的栅格翼展开用时约为 122 ms,略大于本文建立的基于定常流体力求解模型计算结果,计算结果与试验数据分析结果的相对误差约为 -11.7%。

综上,从试验角度进一步验证了本文建立的栅格翼动态展开过程参数预示方法具有工程应用价值。

表 2 航行体水下试验栅格翼展开用时情况

Tab. 2 Expansion time measured in the underwater vehicle experiments

试验序号	展开用时/ms
1	126
2	118
平均值	122

4 结论

本文对水下应用栅格翼动态展开过程进行研究,建立了栅格翼动态展开过程动力学模型,基于定常力计算模型,求解获得栅格翼展开特征角度定常流体力矩,通过对展开过程受力特性的分析拟合,获得展开全程流体力矩随展开角度变化的规律,实现对水下应用栅格翼展开全程动力学模型的求解,获得水下应用栅格翼展开全程参数变化规律,形成了栅格翼动态展开过程参数工程预示方法。通过与水下应用栅格翼非定常流场仿真计算数据以及水下航行体弹射试验数据对比,验证了本文所建立的工程预示方法的正确性及适用性,为水下应用栅格翼方案设计的优化及栅格翼动态强度设计提供了设计参考。

参考文献 (References)

[1] 黎汉华, 石玉红. 栅格翼国内外研究现状及发展趋势[J]. 导弹航天与运载技术, 2008(6): 27-30.
LI HANHUA, SHI YUHONG. Current status and development

trend of grid fin [J]. *Missile and Space Vehicle*, 2008(6): 27 – 30. (in Chinese)

[2] 雷歌, 邓飞, 刘权, 等. 水下航行器折叠翼展开机构设计与动力学仿真[J]. *鱼雷技术*, 2013, 21(2): 81 – 85.
LEI Ge, DENG Fei, LIU Quan, et al. Design and dynamic simulation of folding wing expansion mechanism for underwater vehicle [J]. *Torpedo Technology*, 2013, 21(2): 81 – 85. (in Chinese)

[3] 黄涛, 吴磊, 鲁传敬, 等. 栅格翼和雷体组合体的空泡水动力计算与分析[J]. *水动力学研究与进展*, 2006, 21(2): 239 – 243.
HUANG Tao, WU Lei, LU Chuanjing, et al. Cavitaing grid gin hydrodynamics for missile applications [J]. *Journal of Hydrodynamics*, 2006, 21(2): 239 – 243. (in Chinese)

[4] 朱国祥, 韩茹宗, 罗金玲, 等. 飞航导弹栅格翼数值与试验研究[C]//第一届近代试验空气动力学会议论文集, 2007: 351 – 354.
ZHU Guoxiang, HAN Ruzong, LUO Jinling, et al. Numerical simulation and wind tunnel investigation for grid fin of winged missiles [C]//Proceedings of the First Modern Experimental Aerodynamics Conference, 2007: 351 – 354. (in Chinese)

[5] 李晓晖, 李怀念, 吴俊全. 火箭弹折叠翼展开过程的计算与试验研究[J]. *航天器环境工程*, 2009, 26(12): 82 – 84.
LI Xiaohui, LI Huainian, WU Junquan. Calculation and experimental research on the expansion of rocket folding fin[J]. *Spacecraft Environment Engineering*, 2009, 26(12): 82 – 84. (in Chinese)

[6] 傅慧萍, 李杰. 附加质量 CFD 计算方法研究[J]. *哈尔滨工程大学学报*, 2011, 32(2): 148 – 152.
FU Huiping, LI Jie. Numerical studies of added mass based on the CFD method [J]. *Journal of Harbin Engineering University*, 2011, 32(2): 148 – 152. (in Chinese)

[7] 吴晓军, 马明生, 邓有奇, 等. 结构/非结构混合网格数值模拟栅格翼[J]. *空气动力学学报*, 2009, 27(4): 419 – 424.
WU Xiaojun, MA Mingsheng, DENG Youqi, et al. Navier-Stokes computations of a grid fin missile on hybrid structured-unstructured grids [J]. *ACTA Aerodynamica Sinica*, 2009, 27(4): 419 – 424. (in Chinese)

[8] 俞建阳. 带栅格翼的水下航行体三维流场数值模拟[D]. 哈尔滨: 哈尔滨工业大学, 2012.
YU Jianyang. Three-dimensional numerical simulation of underwater vehicle with grid fins [D]. Harbin: Harbin Institute of Technology, 2012. (in Chinese)

[9] 姚琰, 毛鸿羽. 栅格翼流体动力性能数值模拟[J]. *战术导弹技术*, 2004(2): 13 – 17.
YAO Yan, MAO Hongyu. Numerical simulation of hydrodynamic characteristics for grid fins [J]. *Tactical Missile Technology*, 2004(2): 13 – 17. (in Chinese)

[10] 黄涛. 栅格翼空泡流的数值模拟计算[D]. 上海: 上海交通大学, 2004.
HUANG Tao. Numerical simulations of the cavity flow in grid fin [M]. Shanghai: Shanghai Jiao Tong University, 2004. (in Chinese)

无限质量降落伞充气动力学数值模拟*

高兴龙^{1,2}, 张青斌¹, 高庆玉¹, 唐乾刚¹
(1. 国防科技大学 航天科学与工程学院, 湖南 长沙 410073;
2. 中国空气动力研究与发展中心 设备设计及测试技术研究所, 四川 绵阳 621000)

摘要:为分析降落伞火星再入环境下的超声速开伞性能,基于任意欧拉-拉格朗日罚函数法和多介质任意拉格朗日欧拉算法,求解可压缩流场与降落伞结构的耦合动力学模型。数值模拟盘缝带伞超声速开伞过程外形变化,预测气动力作用下的伞衣织物三维结构动力学行为。结合风洞试验数据,对比分析降落伞开伞性能和前置体对伞衣充气外形的影响。最终给出超声速伞周围非稳态流场的尾流和激波分布。仿真结果表明:盘缝带伞在超声速开伞过程中被完全充满且充气效果良好,未出现塌陷情况;随着来流马赫数的增加,降落伞阻力系数逐渐减小,充气时间缩短。仿真结果与试验结果保持一致,验证了所提方法的有效性。

关键词:降落伞;超声速流动;无限质量充气;流固耦合;可压缩流

中图分类号:V441.8 **文献标志码:**A **文章编号:**1001-2486(2017)03-058-06

Numerical simulation on parachute's infinite mass inflation dynamics

GAO Xinglong^{1,2}, ZHANG Qingbin¹, GAO Qingyu¹, TANG Qiangang¹
(1. College of Aerospace Science and Technology, National University of Defense Technology, Changsha 410073, China;
2. Facility Design and Instrumentation Institute of China Aerodynamics Research and Development Center, Mianyang 621000, China)

Abstract: To analyze the supersonic opening performance of the parachute in Mars reentry environment, the coupling dynamic models between compressible fluid and flexible structure of parachute were solved on the basis of the arbitrary Euler-Lagrange penalty function method and the multi-material arbitrary Lagrange Euler algorithm. The evolution of 3D shape of DGB (disk gap band) parachute during supersonic inflation was simulated, and the structural dynamic behaviors of canopy fabric were predicted. The drag area and coefficients were compared with the wind tunnel data, and the inflation performance of parachute and the influence of fore-body were analyzed. Finally, the wake of unsteady fluid and distribution of shock wave around supersonic parachute were investigated. The results show that: the DGB parachute is well inflated without serious collapse; as the increase of Mach numbers, the drag coefficients gradually decrease, along with the increase of the inflation time, which brings into correspondence with the test results, and proves the validity of the proposed method.

Key words: parachute; supersonic flow; infinite mass inflation; fluid structure interaction; compressible fluid

进入、减速和着陆(Entry, Descent, and Landing, EDL)技术是深空探测实施过程的关键技术之一,而降落伞作为EDL技术的重要组成部分,是火星探测器实现软着陆的关键环节。火星表面大气稀薄,降落伞减速工作处于低密度、超音速、低动压的工作环境^[1],这些特点令火星探测中的降落伞开伞过程变得更为复杂。

降落伞充气过程涉及伞衣结构与气动压力的相互作用,其流固耦合过程的求解作为降落伞研究领域的难点问题一直备受人们关注^[2]。对于开缝形式的复杂伞衣结构,开伞性能更难以准确预测^[3-5]。

降落伞工作过程的流固耦合性能可以借助数值模拟技术进行预测。Lingard等采用任意拉格朗日欧拉(Arbitrary Lagrangian Euler, ALE)方法数值模拟了超声速充气的过程,分析了前体尾流作用下的降落伞阻力性能,以及拖曳比和来流马赫数对伞衣阻力系数的影响^[6-7],但降落伞初始构型为半张满状态,未考虑初始充气过程的影响。Karagiozis等采用大涡模拟的方法,结合自适应网格重构技术进行了可压缩流场与盘缝带(Disk Gap Band, DGB)伞的流固耦合仿真,尤其对流场的湍流特性和伞衣喘振现象进行了很好的模拟^[8]。但该研究只是分析了稳态过程,即伞衣初

* 收稿日期:2016-01-16
基金项目:国家自然科学基金资助项目(11272345,51375486);国防科技大学基金资助项目(JC13-01-04)
作者简介:高兴龙(1987—),男,吉林蛟河人,博士研究生,E-mail:18674853560@163.com;
张青斌(通信作者),男,副教授,博士,硕士生导师,E-mail:qingbinzhang@sina.com

始构型为充满状态。

国内对超声速降落伞的研究起步较晚,荣伟等针对火星探测任务的可行性分析开展了一系列盘缝带伞空投试验和稀薄大气环境的降落伞减速技术研究^[9-10]。彭勇和张青斌等对返回着陆的大型降落伞充气过程进行了分阶段研究,并对伞绳拉直和流固耦合特性进行了分析^[11-12]。目前针对火星环境的降落伞充气过程三维数值模拟和动力学特性的研究比较缺乏。

1 盘缝带伞系统模型

盘缝带伞顶部主盘为原型伞盖,中间开有伞顶孔用于缓解气动压力。主盘与织物带之间开有一条裂缝,充当通气口,主要防止高过载情况下伞衣织物的破裂。在地球对模拟火星环境进行该伞形的测试试验难度很大,尤其对于大载荷的全尺寸伞形,花费巨大。本研究所用的降落伞采用美国国家航空航天局(National Aeronautics and Space Administration, NASA)的火星科学实验室(Mars Science Laboratory, MSL)探测任务所用到的盘缝带全尺寸伞模型(如图 1 所示)。该盘缝带伞与经典的 Viking 伞类似,名义直径 $D_0 = 21.35 \text{ m}$,是目前火星探测任务中所用到的最大的降落伞。探测器为 70° 半锥角的球锥,直径 $D_B = 4.5 \text{ m}$,载荷重量接近 900 kg 。降落伞与探测器之间通过吊带连接,定义拖曳比为 x/D_B ,其中 x 为伞衣底边外缘与球锥最宽处的垂直距离。为保证伞衣开伞过程相对稳定,根据文献[13]结果选取较为安全的拖曳比 $x/D_B = 14$ 。

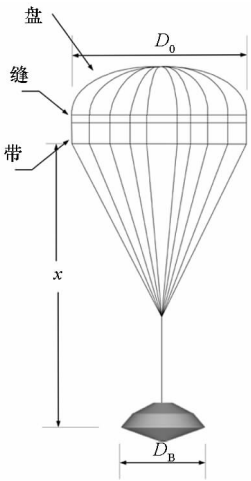


图 1 盘缝带伞-探测器模型

Fig. 1 Model of DGB parachute and probe

该盘缝带伞形为 Viking 型,根据降落伞名义直径可以确定伞衣幅的相关几何参数,伞衣幅及前置体的其他几何参数可以参考文献[14]。

2 控制方程

2.1 流固耦合方程

ALE 流固耦合方法可以求解自由界面流动以及典型的流固耦合问题,具有物质输运功能的单元网格可以在欧拉体系和拉格朗日体系的网格间平动,且在耦合交界面处流体节点随结构变形运动。将流体质点速度通过质量和动量守恒定律在可压流体域内进行离散,即:

$$\mathbf{v}_{i,i} = 0 \text{ in } \Omega_F \times [0, t] \quad (1)$$

$$\frac{\partial \mathbf{v}_i}{\partial t} + (\mathbf{v}_j - \mathbf{v}_j^m) \mathbf{v}_{i,j} - \frac{1}{\rho_F} \boldsymbol{\tau}_{ij} = \mathbf{g}_i \text{ in } \Omega_F \times [0, t] \quad (2)$$

其中: \mathbf{v}_i 为流体速度; ρ_F 为流体密度; \mathbf{v}_j^m 为网格移动速度; Ω_F 为时间域。如果 $\mathbf{v}_j^m = \mathbf{0}$, 则得到欧拉算式, 网格对流速度为空; 如果 $\mathbf{v}_j^m = \mathbf{v}_j$, 则得到拉格朗日算式, 对流速度即为流体速度。 $\mathbf{v}_j - \mathbf{v}_j^m$ 为相对速度, 应力张量 $\boldsymbol{\tau}_{ij}$ 通常定义为:

$$\boldsymbol{\tau}_{ij} = \mu_F (\mathbf{v}_{i,j} + \mathbf{v}_{j,i}) - P \delta_{ij} \quad (3)$$

式中: μ_F 为流体动力黏性系数, P 为压力, δ_{ij} 为狄拉克函数。

给出流场初始和边界条件, 可以对流体动量方程进行求解, 即:

$$\mathbf{v}_i(0) = \mathbf{0} \text{ in } \Omega_F \quad (4)$$

$$\mathbf{v}_i = \hat{\mathbf{v}}_i \text{ on } \delta\Omega_{DF} \times [0, t] \quad (5)$$

式中: $\hat{\mathbf{v}}_i$ 为在流体边界 $\delta\Omega_{DF}$ 施加的速度集合。这里可以将流场有限元模型的底部单元设置为压力入口单元集合, 并对入口单元施加速度载荷。

对于低密度气体, 可以采用理想气体状态方程进行模拟, 给定初始压力和初始内能, 对能量控制方程进行求解, 即:

$$\begin{cases} P = \rho(C_p - C_v)T \\ C_p = C_{p0} + C_L T + C_Q T^2 \\ C_v = C_{v0} + C_L T + C_Q T^2 \end{cases} \quad (6)$$

式中, C_p 和 C_v 分别为定压和定容下的比热容, ρ 为大气密度, C_L 和 C_Q 分别为温度对应系数, T 为温度。火星环境大气参数的比热容 γ 为 1.29, 大气压强为 750 Pa 。

2.2 材料本构模型

伞衣材料模型选择柔性织物材料模型, 该材料可适用于薄膜单元, 具有非线性动力学特性且能够承受大变形行为, 可以模拟伞衣薄膜材料。伞绳采用自定义的非线性本构模型:

$$F = \begin{cases} 0 & \varepsilon \leq 0 \\ p(\varepsilon) + C \cdot \dot{\varepsilon} & \varepsilon > 0 \end{cases} \quad (7)$$

式中: $p(\varepsilon)$ 为伞绳的非线性张力函数; C 为阻尼系数;应变 ε 为

$$\varepsilon = \frac{\Delta l}{l_0 - l_{\text{off}}}$$

(8)

式中, Δl 为伞绳长度变化, l_0 为伞绳初始长度, l_{off} 为初始长度偏移量。

2.3 耦合界面

选择显示动力学积分方法求解充气过程的流固耦合问题,程序在每个时间步内首先需要分别计算流场网格和伞衣结构网格的节点力,之后采用罚函数法将流体-结构交界面的节点力进行耦合。假设在 $t=t^n$ 时刻,主节点(结构节点)与从节点(流体节点)之间的穿透深度为 d^n , v_{rel} 为主从节点相对速度,则可对 d^n 进行迭代更新,即:

$$d^{n+1} = d^n + v_{\text{rel}}^{1+n/2} \cdot \Delta t$$

(9)

降落伞伞衣为柔性透气性织物,对于渗透介质的耦合力可以通过 Shell 单元体积的 Ergun 方程导出^[15]:

$$\frac{dP}{dr} = a(\mu, \varepsilon) \cdot v_{\text{rel}} + b(\rho, \varepsilon) \cdot v_{\text{rel}}^2$$

(10)

式中: r 为壳单元的法向; $a(\mu, \varepsilon)$ 为渗透性壳单元的渗透系数; $b(\rho, \varepsilon)$ 为惯性系数, a 、 b 系数组合即为伞衣透气性参数。

3 数值仿真

3.1 有限元建模

本文主要模拟降落伞从伞包中拉直后并开始充气直至充满稳定的过程,因此降落伞初始状态设计为折叠状态。如图 2 所示,伞衣外形为锥形,气流自低端开口处流入。



图 2 盘缝带伞初始折叠有限元模型
Fig.2 Finite element model of initial folded disk-gap parachute

本文是模拟风洞试验的无限质量充气情况,根据仿真经验及文献^[7]给出的流场域尺寸设置参考建立圆柱形流场。流场几何外形为圆柱形,降落伞系统沿高度方向置于中轴线位置。圆柱直径为 $4D_0$,顶部距离伞衣高度为 $5D_0$,底部与再入体距离 $2D_B$ 。

降落伞结构有限元模型主要分为伞衣、伞绳

部件,分别采用薄膜单元和离散梁单元进行网格划分。前置体假设为刚体,直接采用实体单元进行划分。流场为空气介质,采用六面体实体单元进行网格划分,靠近伞衣附近网格进行局部加密,有限元模型统计信息见表 1。

表 1 流固耦合有限元模型统计信息
Tab.1 Summary of finite element model information for fluid-solid coupling

名称	节点	单元	类型
伞衣	12 960	12 348	2DShell
伞绳	3659	5904	1DLink
前置体	1487	236	3DSolid
流场	645 567	625 684	3DSolid

3.2 仿真工况

暂不考虑伞绳的流固耦合效应,仅设置伞衣薄膜单元与流场单元的耦合接触。分析开伞动压对开伞充气过程的影响。雷诺数变化范围为 $7 \times 10^6 \sim 1.3 \times 10^7$ 。具体仿真工况参数见表 2。

表 2 仿真工况参数^[16]
Tab.2 Parameters of simulations^[16]

马赫数	雷诺数 ($\times 10^6$)	开伞动压/ kPa	来流速度/ (m/s)
1.5	7.4	22.5	436.6
2.0	9.6	29.3	524.4
2.5	12.4	34.7	589.2

流场对流算法选择二阶精度 Van Leer MUSCL 格式,相比于程序中的一阶精度 donor cell 格式,该格式结果更稳定且计算成本较低,能够减少能量耗散,更好地模拟激波和尾流。

4 计算结果及分析

4.1 降落伞外形变化

图 3 为马赫数为 2.0 时充气过程中盘缝带伞的外形变化。从图中可以看到,伞衣首先自低端充气,并被逐渐拉直为管状,该阶段称之为“初始充气”。之后气流在伞衣顶部积聚,“主充气阶段”开始,气流自顶端沿伞衣径向展开,伞衣呈“灯泡状”,之后伞衣逐渐充满。可以看出,数值仿真得到的结果基本重现了实物伞开伞过程,外形变化符合降落伞开伞过程外形变化规律。尤其是伞衣初始充气所出现的“灯泡状”外形也得到

了较好的数值模拟,伞衣最终张满状态稳定。

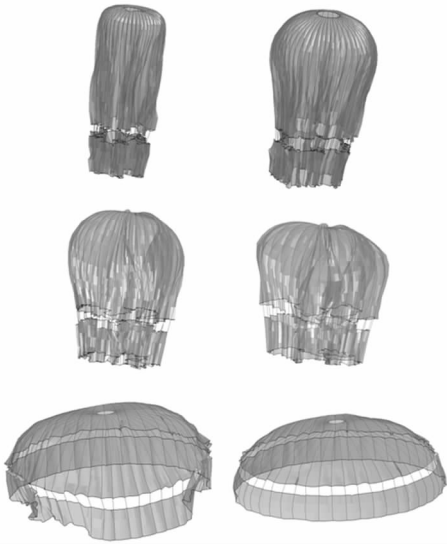


图 3 降落伞充气过程外形变化

Fig.3 Evolution of parachute shape during inflation

图 4 为盘缝带伞充满状态的仿真结果(左图)与 MSL 试验测试伞(右图)的外形对比。单独采用绝对尺寸的伞衣投影直径和伞衣高度无法准确地对比伞衣外形大小,通常采用伞衣高与投影直径的比值进行对比分析。对于 MSL 试验的 DGB 伞,该比值接近 0.5,本文仿真计算的测量结果约为 0.48,与试验值较为接近。同时从图中可以观察到,仿真模型与试验伞在底部伞带区域略有不同,试验伞的伞带“鼓包”现象更为明显,这主要跟伞衣材料模型有关,其无法完全真实地模拟柔性织物结构的张力状态。

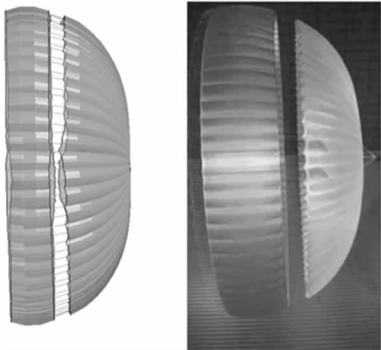


图 4 盘缝带伞充满状态模型与 MSL 试验模型侧视图对比^[17]

Fig.4 Profile views of inflated model of simulation and MSL tested DGB parachute^[17]

图 5 为不同来流动压下的伞衣阻力面积变化。从图中可以看出,伞衣充满后发生明显波动,超声速伞的呼吸现象对降落伞的稳定性会产生显著影响,甚至会出现伞衣局部塌陷情况。但最终

伞衣仅出现小幅度的喘振并维持在某一频率范围内,趋于相对稳定状态。同时随着马赫数和开伞动压的增加,开伞时间缩短,伞衣充满外形并无明显变化。

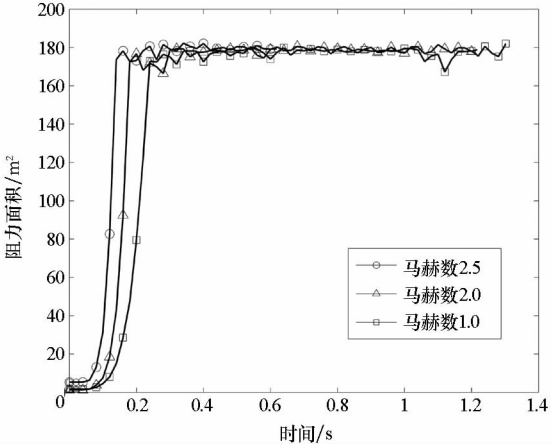


图 5 开伞过程阻力面积变化对比

Fig.5 Evolution of drag area during inflation

4.2 开伞力与阻力系数

图 6 为开伞力变化曲线与试验数据的对比,可以看出,在充气初始时刻出现一定的伞绳回弹现象。之后开伞力逐渐增加,充气过程开始,直至首次出现开伞力峰值,此时伞衣也完全充满。之后开伞力明显回落,并保持不变。仿真结果与试验数据结果趋势一致,仿真计算的开伞力峰值为 364.7 kN,略大于试验数据的 352.6 kN。

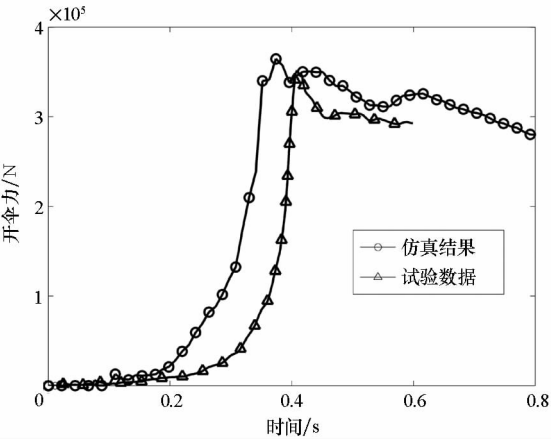


图 6 开伞力变化对比^[18]

Fig.6 Comparison of opening forces^[18]

同时可以计算不同工况下的阻力系数:

$$C_d = \frac{F_D}{qS_d} \tag{11}$$

式中, C_d 为阻力系数, F_D 为开伞力, q 为来流动压, S_d 为伞衣阻力面积。图 7 为阻力系数随马赫数的变化结果及其与试验结果的对比,可以看出数值模拟结果与试验结果趋势相符,即随着马赫数的

增加,阻力系数下降,同时数值模拟结果整体要比实际结果偏高。

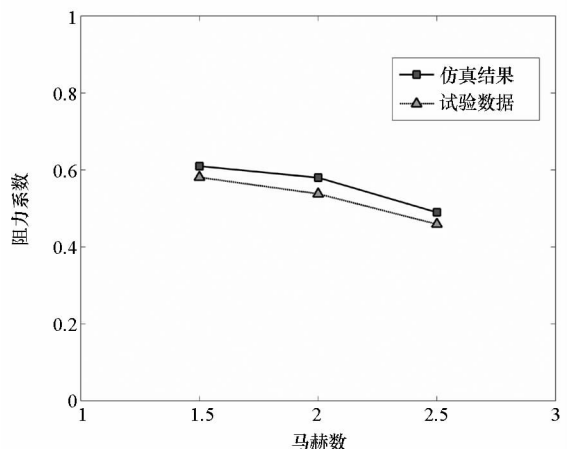


图 7 稳态阻力系数对比^[19]

Fig. 7 Comparison of steady drag coefficients^[19]

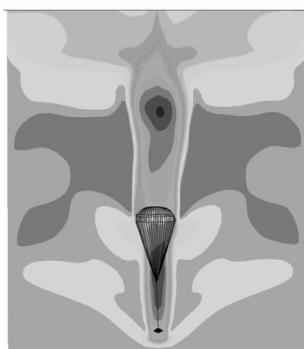
4.3 流固耦合结果

图 8 为充气过程伞衣周围流场速度矢量云图。EDL 过程中,降落伞在超声速流中运动开伞,伞衣形成半球状阻流体,会在稀薄大气中压缩周围气流形成激波。前置体的存在会明显改变降落伞周围流场分布,球锥外形会在超声速流中形



(a) 充气过程来流速度分布

(a) Velocity distribution during inflation



(b) 伞衣张满时流场速度分布

(b) Velocity distribution of inflated parachute

图 8 充气过程流场速度矢量云图

Fig. 8 Velocity contour of fluid during inflation

成前体激波,而前体尾流则会形成低速非稳态紊流区域,与伞衣前体弓波发生耦合,会在来流方向呈现极不稳定的流场分布。从图中可以看出,伞衣充气过程来流速度非对称分布,伞衣在尾流处摆动,且出现局部塌陷情况,但最终伞衣张满,流场速度实现对称分布。

通过流固耦合计算,可以同时得出超声速流中伞衣的结构动力学响应,图 9 为开伞后伞衣充满状态的 Von Mises 应力分布。从图中可以明显看出伞顶孔附近区域应力较为集中,且结构变形的平均水平要高于其他区域,因此在对伞衣进行强度校核时应选取该处区域的应力应变水平作为参考。

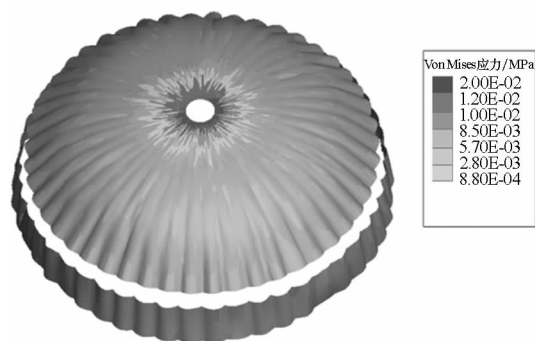


图 9 伞衣表面应力分布云图

Fig. 9 Stress distribution contour on canopy

5 结论

基于 ALE 罚函数耦合方法对盘缝带伞火星环境的无限质量充气过程进行了数值模拟。流场采用可压缩流进行求解,边界条件选取为常值压力入口边界以模拟风洞吹风的效果。结果表明:充气过程伞衣在短暂时间内呈“灯泡状”,之后伞衣被拉为柱形,在伞顶内部气流扩散作用下伞衣底边迅速膨胀并充满,伞衣幅呈明显的“鼓包”现象。同时随着来流动压的增加,开伞时间减少,阻力系数减小。前置体对降落伞开伞效果影响明显,尾流容易与降落伞前体激波产生耦合效应,使得流场在来流方向不对称分布,影响降落伞稳定性。最终仿真结果与试验数据一致,验证了本文方法的有效性。

参考文献 (References)

- [1] 荣伟, 陈国良. 火星探测器减速着陆技术特点 [J]. 航天返回与遥感, 2010, 31(4): 1-6.
RONG Wei, CHEN Guoliang. The characters of deceleration and landing technology on Mars explorer [J]. Spacecraft Recovery & Remote Sensing, 2010, 31(4): 1-6. (in Chinese)

[2] 王利荣. 降落伞理论与应用 [M]. 北京: 宇航出版社, 1997.
WANG Lirong. Parachute theory and application [M]. Beijing: China Astronautic Publishing House, 1997. (in Chinese)

[3] 高兴龙, 唐乾刚, 张青斌, 等. 开缝伞充气过程流固耦合数值研究 [J]. 航空学报, 2013, 34 (10): 2265 – 2276.
GAO Xinglong, TANG Qiangang, ZHANG Qingbin, et al. Numerical study on fluid-structure interaction of slot parachute's inflation process [J]. Acta Aeronautica et Astronautica Sinica, 2013, 34 (10): 2265 – 2276. (in Chinese)

[4] Gao X L, Zhang Q B, Tang Q G, et al. Fluid-structure interaction simulation of parachute in low speed airdrop [C]// Proceedings of the World Congress on Engineering & Computer Science, 2013.

[5] Gao X L, Zhang Q B, Tang Q G. Transient dynamic modeling and analysis of complex parachute inflation with fixed payload [J]. Journal of Aerospace Engineering, 2014: 04014097.

[6] Lingard J S, Darley M G. Simulation of parachute fluid structure interaction in supersonic flow [C]//Proceedings of 18th AIAA Aerodynamic Decelerator Systems Technology Conference and Seminar, AIAA – 2005 – 1607, 2005.

[7] Lingard J S, Darley M, Underwood J C. Simulation of Mars supersonic parachute performance and dynamics [C]// Proceedings of 19th AIAA Aerodynamic Decelerator Technology Conference and Seminar, AIAA – 2007 – 2507, 2007.

[8] Karagiozis K, Kamakoti R, Cirak F. A computational study of supersonic disk-gap-band parachutes using large-eddy simulation coupled to a structural membrane [J]. Journal of Fluids and Structures, 2011, 27 (2): 175 – 192.

[9] 荣伟, 陈旭, 陈国良. 大气密度对降落伞充气性能的影响 [J]. 航天返回与遥感, 2006, 27 (3): 11 – 16.
RONG Wei, CHEN Xu, CHEN Guoliang. The effect of atmospheric density on parachute inflation performances [J]. Spacecraft Recovery & Remote Sensing, 2006, 27 (3): 11 – 16. (in Chinese)

[10] 荣伟, 陈旭, 陈国良. 低密度大气中降落伞开伞动载的研究 [J]. 航天返回与遥感, 2006, 27 (4): 7 – 11.
RONG Wei, CHEN Xu, CHEN Guoliang. The study of the parachute opening load in low atmospheric density [J]. Spacecraft Recovery & Remote Sensing, 2006, 27 (4): 7 – 11. (in Chinese)

[11] 张青斌, 程文科, 彭勇, 等. 降落伞拉直过程的多刚体模型 [J]. 中国空间科学技术, 2003, V23 (2): 45 – 50.
ZHANG Qingbin, CHENG Wenke, PENG Yong, et al. A multi-rigid-body model of parachute deployment [J]. Chinese Space Science and Technology, 2003, V23 (2): 45 – 50. (in Chinese)

[12] 彭勇, 张青斌, 秦子增. 降落伞主充气阶段数值模拟 [J]. 国防科技大学学报, 2004, 26 (2): 13 – 16.
PENG Yong, ZHANG Qingbin, QIN Zizeng. Simulation of parachute final inflation phase [J]. Journal of National University of Defense Technology, 2004, 26 (2): 13 – 16. (in Chinese)

[13] Cruz J R, Mineck R E, Keller D F, et al. Wind tunnel testing of various disk-gap-band parachutes [C]// Proceedings of 17th AIAA Aerodynamic Decelerator Systems Technology Conference and Seminar, AIAA – 200 – 2129, 2003.

[14] Cruz J R, David W, Jeremy S, et al. Parachute models used in the Mars science laboratory entry, descent, and landing simulation [C]//Proceedings of AIAA Aerodynamic Decelerator Systems (ADS) Conference, AIAA – 2013 – 1276, 2013.

[15] Ergun S. Fluid flow through packed beds [J]. Chemical Engineering Progress, 1952, 48 (2): 89 – 94.

[16] engupta A, Roeder J, Kelsch R, et al. Supersonic disk gap band parachute performance in the wake of a Viking-type entry vehicle from Mach 2 to 2. 5 [C]//Proceedings of AIAA Atmospheric Flight Mechanics Conference and Exhibit, AIAA – 2008 – 6217, 2008.

[17] Witkowski A, Kandis M, Sengupta A, et al. Comparison of subscale versus full-scale wind tunnel tests of MSL disk gap band parachutes [C]//Proceedings of 20th AIAA Aerodynamic Decelerator Systems Technology Conference and Seminar, AIAA – 2009 – 2914, 2009.

[18] Witkowski A, Kandis M, Adams D S. Inflation characteristics of the MSL disk gap band parachute [C]//Proceedings of 20th AIAA Aerodynamic Decelerator Systems Technology Conference and Seminar, AIAA – 2009 – 2915, 2009.

[19] Sengupta A, Steltzner A, Comeaux K. Results from the Mars science laboratory parachute decelerator system supersonic qualification program [C]//Proceedings of 2008 IEEE Aerospace Conference, 2008: 1 – 15.

微处理器容软错误设计量化评估指标及评估方法*

龚 锐,郭御风,邓 宇,石 伟,窦 强
(国防科技大学 计算机学院,湖南 长沙 410073)

摘 要:针对高可靠微处理器软容错设计,提出了一种新的可靠性度量标准,增强的平均无失效工作量,以解决现有度量标准没有综合考虑性能、面积、功耗开销带来的可靠性降低的缺点;提出了一种评估方法对增强的平均无失效工作量以及两种控制流检测技术进行定量评估。评估结果表明,软硬件结合的控制流检测技术较好地折中了可靠性、性能、面积和功耗。量化评估指标全面考虑了多种开销对微处理器可靠性的影响,采用相应的评估方法可以更加准确地对微处理器可靠性加固手段进行定量评估,以指导设计探索和设计优化。

关键词:容软错误;量化评估;评估方法;微处理器;可靠性

中图分类号:TP302.8 **文献标志码:**A **文章编号:**1001-2486(2017)03-064-05

Quantitative evaluation metric and methodology for microprocessor soft error tolerance design

GONG Rui, GUO Yufeng, DENG Yu, SHI Wei, DOU Qiang

(College of Computer, National University of Defense Technology, Changsha 410073, China)

Abstract: Aiming at highly reliable microprocessor soft error tolerance design, a new metric, eMWTF (enhanced mean work to failure), was proposed to capture the trade-off among reliability, performance, area and power. A quantitative approach for evaluating eMWTF was also presented. Two control flow checking techniques were quantitatively evaluated in reliability. The experimental results indicate that the control flow checking by compiler signatures and hardware checking achieves better trade-off among reliability, performance, area and power. Because the eMWTF metric takes into consideration performance, area and power overheads, the quantitative reliability evaluation can be more accurate by using this metric and corresponding methodology. Finally, the evaluation results can effectively guild the design exploring and optimization.

Key words: soft error tolerance; quantitative evaluation; evaluation methodology; microprocessor; reliability

应用于复杂电磁环境的集成电路受到高能粒子轰击,会发生瞬时充放电,使得逻辑状态发生翻转,这种由高能粒子轰击所引发的错误被称为“软错误”。高可靠微处理器一般采用多种容软错误设计技术。这些容软错误设计在提高微处理器可靠性的同时,不可避免地带来了性能、面积、功耗的开销。最新的软错误发生机理研究表明,性能、面积、功耗的开销对于微处理器的可靠性有负面影响。

1 研究背景

1.1 软错误类型

高能粒子引起的微处理器软错误包括单事件翻转 (Single Event Upset, SEU)、单事件瞬态 (Single Event Transient, SET) 和多位翻转 (Multi

Bit Upsets, MBU) 等。其中 SEU 是指单个存储单元遭到高能粒子轰击而发生的逻辑翻转。翻转后错误的值将一直被保持到下一次写入操作。SET 是指高能粒子轰击导致组合逻辑通路上产生的毛刺。这种 SET 毛刺有可能沿组合逻辑通路传递,也可能被电路自身的结构所屏蔽。当 SET 毛刺恰好在时钟沿传递到时序逻辑输入,错误的值将会被采样,导致微处理器功能错误。此外,随着集成电路特征尺寸的缩小和集成度的提高,一次粒子轰击有可能导致动态随机访问存储器 (Dynamic Random Access Memory, DRAM) 存储阵列或静态随机访问存储器 (Static Random Access Memory, SRAM) 存储阵列内相邻的多个存储单元发生翻转,这种类型的软错误被称为 MBU。

* 收稿日期:2015-11-13
基金项目:国家自然科学基金资助项目(61202123,61202122,61402497)
作者简介:龚锐(1980—),男,四川雅安人,助理研究员,博士,E-mail:rgong@nudt.edu.cn

与设计制造过程中引入的硬错误相比,上述软错误具有瞬态、可恢复、发生位置和时间随机等特点。

1.2 软错误发生机理

电子器件发生软错误的概率受辐射水平、存储电荷及敏感源漏区域面积的影响。一般采用软错误率(Soft Error Rate, SER)^[1]来表征器件发生软错误的概率。SER可以采用式(1)推算^[2]。

$$SER \propto F \cdot A_{sd} \cdot \exp\left(-\frac{Q_{crit}}{Q_s}\right) \quad (1)$$

其中: F 是能量大于1 MeV的高能粒子流密度; A_{sd} 是对辐射敏感的面积,对单个晶体管器件来说,即源漏极面积; Q_{crit} 是导致芯片中存储信息发生逻辑翻转所需要的最小电量,称为临界电量^[3-4]; Q_s 则是粒子轰击在芯片上引起的实际充放电电量。

1.3 容软错误能力量化评估

一般来说,对软错误进行检测、屏蔽与恢复,都需要某种冗余机制。这些冗余设计不可避免地带来了芯片面积、程序执行性能和微处理器功耗的开销。微处理器受到粒子轰击的概率正比于其暴露于辐射环境中的芯片面积,芯片面积的增加将导致更多的软错误。程序执行性能的降低,将增加单个程序的执行时间,从而增加单个程序执行过程中受到高能粒子轰击的概率。微处理器功耗的上升将导致芯片工作温度的升高,根据国内外研究人员在电路级的研究,SEU对温度的变化不敏感^[5],但在 $-55 \sim +125$ °C范围内,SET毛刺的宽度随温度的升高而变大,其宽度与温度基本呈线性变化^[6]。SET毛刺的展宽将增加其被下级时序逻辑单元采样到的概率,从而增加微处理器发生软错误的概率。因此,冗余设计带来的面积、性能、功耗的开销对微处理器的容软错误能力是有负面影响的。片面强调容软错误的冗余设计而忽视其开销带来的负面影响,并不一定能获得最优化的可靠性提升。

2 相关工作

可靠性评估中重要的量化评估指标是平均无失效时间(Mean Time To Failure, MTTF),该参数表示微处理器发生失效的期望时间。在容软错误能力评估中,MTTF可以简单表示为:

$$MTTF = \frac{1}{SER} \quad (2)$$

由于相当一部分的软错误会被微处理器体系

结构的固有特性或各种软错误加固技术所屏蔽,并不会引起程序的执行结果发生错误。因此,文献[7]提出了结构弱点因子(Architectural Vulnerability Factor, AVF)来表示原始软错误导致微处理器失效的概率,该参数也可以表征微处理器体系结构所具有的软错误屏蔽能力。在此基础上MTTF可以更精确地表示为:

$$MTTF = \frac{1}{SER \cdot AVF} \quad (3)$$

采用MTTF进行可靠性评估,只考虑了容软错误技术带来的可靠性提升(即AVF的降低),而未考虑其面积、性能、功耗开销带来的可靠性降低。文献[8]给出了平均无失效指令(Mean Instruction To Failure, MITF)的概念。MITF表征微处理器在失效前可以执行的平均指令条数,可以表示为:

$$MITF = IPC \cdot Frequency \cdot MTTF = \frac{IPC \cdot Frequency}{SER \cdot AVF} \quad (4)$$

其中,IPC表示每周可执行的指令条数,Frequency表示微处理器频率。

文献[9]进一步推广,提出了平均无失效工作量(Mean Work To Failure, MWTF)的概念来表征微处理器在失效前可完成的平均工作量。MWTF定义为:

$$MWTF = \frac{1}{SER \cdot AVF \cdot t_{exe}} \quad (5)$$

其中, t_{exe} 为微处理器执行给定工作所需的时间,一般表示为执行一组测试程序所需的时间。

MITF和MWTF两个量化指标考虑了性能开销对微处理器容软错误能力的影响,但仍未考虑面积和功耗的影响。在前期的研究中,提出了改进的平均无失效工作量(modified MWTF, mMWTF)的概念^[10],将面积和性能开销都纳入量化评估指标内。mMWTF定义为:

$$mMWTF = \frac{1}{SER \cdot A \cdot AVF \cdot t_{exe}} \quad (6)$$

其中A表示芯片面积。

上述相关工作一步步推进可靠性量化评估向更全面的方向发展,但仍未将功耗因素考虑在内。

3 量化评估指标

已有的可靠性量化评估指标中,一般采用SER来表征微处理器在单位时间内发生的软错误。由于芯片面积不同,辐射面积就不相同,SER也不同。精确定义瞬态故障率(Transient Fault Rate, TFR)为单位芯片面积微处理器在单位时间

内发生 SEU、SET 等瞬态故障的概率。可以认为,在相同的制造工艺和相同的辐射条件下,微处理器的 TFR 相同。

现只考虑 SEU 和 SET 两种类型的瞬态故障。定义 AVF 为 SEU 导致微处理器发生失效的概率。定义 TVF 为 SET 被寄存器采样而发生 SEU 的概率。AVF 表征了体系结构和软件对 SEU 的屏蔽能力,而 TVF 则表征了寄存器采样窗口对 SET 的屏蔽能力。SET 被采样形成 SEU 后,也只有 AVF 导致微处理器发生失效。假设微处理器中发生的 SET 占有瞬态故障的百分比为 P_{SET} ,则在单位时间内微处理器发生失效的次数为:

$$\begin{aligned} N_f &= N_{\text{f_SET}} + N_{\text{f_SEU}} \\ &= TFR \cdot A \cdot P_{\text{SET}} \cdot TVF \cdot AVF + TFR \cdot A \cdot \\ &\quad (1 - P_{\text{SET}}) \cdot AVF \\ &= TFR \cdot A \cdot AVF \cdot [P_{\text{SET}} \cdot TVF + (1 - P_{\text{SET}})] \end{aligned} \quad (7)$$

即单位时间内发失效的次数 N_f 为由 SET 引起的失效数 $N_{\text{f_SET}}$ 和由 SEU 引起的失效数 $N_{\text{f_SEU}}$ 的总和。

对于 SET 来说,其 TVF 可以近似表征为 SET 脉冲宽度 W 与寄存器时钟频率 T_{clk} 的比值,即:

$$TVF = \frac{W}{T_{\text{clk}}} \quad (8)$$

由国内外对软错误机理的研究可知,温度 T 对 SEU 基本没有影响,但会导致 SET 脉冲宽度 W 展宽,且 W 与 T 基本呈线性关系。假设 W 与 T 的关系为:

$$W = aT + b \quad (9)$$

假设 SET 脉冲展宽后仍然小于等于时钟频率 T_{clk} ,那么将式(9)代入式(8),有:

$$TVF = \frac{aT + b}{T_{\text{clk}}} \quad (10)$$

可以简单地认为微处理器工作温度与单位面积功耗(P/A)即功耗密度呈线性关系,所以有:

$$T = x \frac{P}{A} + y \quad (11)$$

将式(11)代入式(10),可得:

$$TVF = \frac{a \left(x \frac{P}{A} + y \right) + b}{T_{\text{clk}}} = \frac{\alpha \frac{P}{A} + \beta}{T_{\text{clk}}} \quad (12)$$

即在时钟频率不变的情况下,由 SET 导致 SEU 的概率 TVF 与单位面积功耗(P/A)呈线性关系。将式(12)代入式(7),有:

$$\begin{aligned} N_f &= TFR \cdot A \cdot AVF \cdot \left[P_{\text{SET}} \cdot \frac{\alpha \frac{P}{A} + \beta}{T_{\text{clk}}} + (1 - P_{\text{SET}}) \right] \\ &\quad (13) \end{aligned}$$

因此,提出增强的平均无失效工作量(enhanced Mean Work To Failure, eMWTF),来表征微处理器在发生失效前可以完成的平均工作量。该量化标准可定义为:

$$eMWTF = \frac{1}{N_f \cdot t_{\text{exe}}} \quad (14)$$

其中, t_{exe} 为完成单位工作量所需的时间,一般表征为完成一组典型测试程序所需的时间。因此 eMWTF 可以表示微处理器在失效前可以完成这种典型测试程序的次数,即可以完成的平均工作量。将式(13)代入式(14),可得:

$$eMWTF = \frac{1}{TFR \cdot A \cdot AVF \cdot \left[P_{\text{SET}} \cdot \frac{\alpha \frac{P}{A} + \beta}{T_{\text{clk}}} + (1 - P_{\text{SET}}) \right] \cdot t_{\text{exe}}} \quad (15)$$

由式(15)可知, eMWTF 是一个涉及了多种设计维度的微处理器可靠性量化评估指标,该指标综合考虑了软错误发生的机理(TFR 和 P_{SET})、容软错误设计带来的可靠性的提升(即 AVF 的降低)以及容软错误设计带来的面积(A)、性能(T_{clk} 和 t_{exe})和功耗(P)开销对可靠性的影响。因此是一个全面准确的量化评估指标。

4 量化评估方法

针对 eMWTF 的量化评估方法如图 1 所示。

该量化评估方法紧密结合半定制的微处理器设计流程。首先在 RTL 级的功能模拟时,执行一组标准的测试程序,获得 t_{exe} 参数。

功能模拟通过后,由综合工具将 RTL 级代码综合为门级网表。在综合的过程中可以知道该设计可以运行的时钟频率 T_{clk} 和所使用的标准单元总面积。由于芯片的总面积 A 需要在最后版图生成后才能确认,但量化评估需要尽可能地在设计的早期进行,所以采用综合时获得的标准单元和 SRAM 总面积信息来近似替代芯片总面积 A 。

通过综合得到门级网表以后,需要在门级网表上进行错误注入模拟,以获得微处理器的 AVF 参数。同时,可以利用前仿时得到的波形信息,通过功耗评估工具获得较准确的功耗参数 P 。

在获得 t_{exe} 、 A 、 T_{clk} 、AVF、 P 等参数以后,可以对 eMWTF 进行量化评估,从而指导设计折中和设计选择。如果没有达到预设的可靠性指标,则需要迭代回去重新进行设计。

需要注意的是,由于 eMWTF 中的某些参数并不能获得准确的数值。如式(15)中的瞬态故障率 TFR 与使用环境的辐照水平相关; SET 故

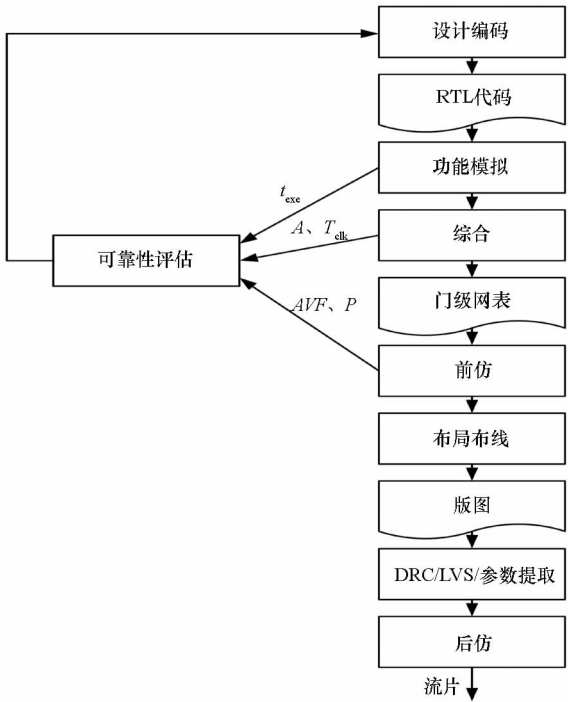


图1 量化评估方法

Fig.1 Quantitative evaluation methodology

障占有所有瞬态故障的百分比 P_{SET} 除了与辐照水平相关,还与芯片内部逻辑相关; TVF 与单位面积功耗(P/A)的线性关系参数 α, β 则与辐照水平、工艺、电气特性等相关。在评估时,只能对这些参数进行假设。此外,在评估的过程中也进行了一些假设,如采用综合时获得的标准单元和 SRAM 总面积来近似替代芯片总面积 A 。因此,采用 eMWTF 和所提出的评估方法,无法获得精确的可靠性数值。但是可以评估出在相同假设条件下,不同可靠性加固手段所能带来的相对的可靠性关系,从而指导设计空间探索和设计选择。

5 量化评估实验及结果

采用 eMWTF 指标对两种控制流检测技术进行可靠性量化评估,并给出相应的结果。

5.1 控制流检测技术

高能粒子导致的故障可能引起控制流错误,即程序的执行流程发生混乱。控制流检测的基本思想是实时监测程序的运行轨迹并与编译预期的轨迹进行比较,以有效防止由于控制流错误导致的系统崩溃。

5.1.1 CFCSS 技术

文献[11]提出了一种纯软件实现的控制流检测 (Control Flow Checking by Software Signatures, CFCSS) 技术。该方法定义程序流图为

有向图 $CFG = (V, E)$, 其中 $V = \{v | v \text{ 为基本块} \}$, $E = \{ \langle v_i, v_j \rangle | \text{存在从 } v_i \text{ 到 } v_j \text{ 的分支或跳转} \}$ 。对于某个特定的基本块 v_i , 赋予其唯一的签名值 S_i 。如果 $\exists \langle v_i, v_j \rangle \in E$, 则 v_i 到 v_j 的签名距离 $d_j = S_i \oplus S_j$, 该签名距离在编译时即可确定。当程序执行从 v_i 到 v_j 的控制流转移时, 计算运行时签名值 $s_j = S_i \oplus d_j$ 。如果分支或转移正确, 则 $s_j = S_i \oplus d_j = S_i \oplus (S_i \oplus S_j) = S_j$ 。如果 $s_j \neq S_j$, 则表明发生了控制流错误。由于采用纯软件实现, CFCSS 比较灵活, 且不用对硬件进行任何改动, 没有额外的面积开销。但是这些签名检测指令若编译为 8051 指令, 执行一次签名检测需要 13 个时钟周期, 性能开销比较大。

5.1.2 CFCCH 方法

为了解决 CFCSS 技术性能开销大的缺点, 文献[12]中提出了一种编译签名硬件检测的控制流检测 (Control Flow Checking by Compiler signatures and Hardware checking, CFCCH) 方法。该方法采用与 CFCSS 技术相同的签名算法, 但只在每个基本块的头部依次插入三个字节的签名数据, 即签名距离 d_i 、签名值 S_i 和运行时调整签名 D_i 。为了实现硬件检测, 增加了两个特殊寄存器 Sreg 和 Dreg, 分别记录当前基本块的签名值并运行时调整签名。在每次控制流转移, 即分支或跳转指令之后, 硬件自动进行一次检测, 若检测无误, 才运行新基本块的指令。每次检测只需要 3 个时钟周期。CFCCH 方法采用硬件进行检测, 有额外的面积开销, 但是性能开销大大减少。

5.2 评估结果

上述 CFCSS 和 CFCCH 两种控制流检测技术各有优劣。分别采用 MTF、MWTF、mMWTF 和 eMWTF 4 种量化评估指标对这两种容软错误技术进行评估, 并且对未经加固设计的 8051 也进行量化评估, 以获得两种加固技术相对于未加固芯片的归一化可靠性参数, 从而指导设计选择。

循环运行测试程序集, 并注入了 10 000 个故障, 以使结果具有统计意义。同时, 采用 65 nm 工艺对 3 款微控制器进行了综合, 约束的时钟频率均为 100 MHz。从综合得到的总的标准单元和 SRAM 面积来看, CFCSS 由于没有任何硬件改动, 没有额外的面积开销。CFCCH 采用了硬件检测, 总的标准单元面积比未采用容软错误技术的非容错 (NO n Fault Tolerance, NFOT) 版本大 7.5%。此外, 采用功耗评估工具对 3 款微控制器进行功耗评估。结果表明, CFCSS 与 NOFT 功耗相当, 但 CFCCH 比 NOFT 约增加了 10.3% 的功耗。运行

了一组测试程序,以便获得性能参数。结果表明,CFCSS 的签名检测代码执行一次需要 13 个周期,性能开销较大,其性能开销为 NOFT 的 54% ~ 112%。CFCCH 在每个检测点只增加了额外的 3 个时钟周期,带来了 9% ~ 36% 的性能开销,低于 CFCSS。

在获得上述 AVF、面积、功耗、性能参数的基础上,为了获得 eMWTF 数值,做出如下假设。假设 P_{SET} 为 0.5,即发生 SET 故障的概率和 SEU 故障概率相同。假设线性关系参数 $\alpha = \beta = 1$ 。在上述假设基础上,获得的评估结果如图 2 所示。

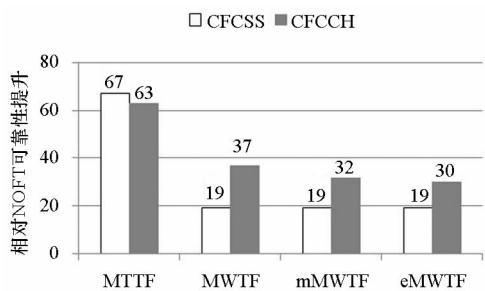


图 2 归一化可靠性评估结果

Fig. 2 Normalized reliability evaluation results

图 2 中的可靠性数值都是与 NOFT 进行了归一化后的相对值。从图 2 中可以看出,由于具有性能开销,同一容软错误技术的 MWTF 值要小于其 MTTF 值。MWTF 表示两次失效之间能够执行的平均工作量,而工作量的定义与实际应用相关。这说明对实际的应用来说,尽管容软错误技术使得两次失效之间能够正常执行的时间大大增加,但是这段时间内所能执行的有用工作量并没有成比例增加。同样地,具有面积开销的容软错误技术,其 mMWTF 值也要小于 MWTF 值。这是因为面积的开销将导致更多的原始软错误。此外,具有功耗开销的容软错误技术,其 eMWTF 值也要小于 mMWTF 值,这是因为功耗开销将导致芯片温度的上升,从而引起 SET 脉冲宽度变大,使其更容易引发寄存器翻转。从以上分析可知,对容软错误技术进行评估时必须全面、定量地考虑性能、面积和功耗的开销,以便进行更好的折中。

6 结论

为解决原有微处理器容软错误评估中不考虑功耗开销的缺点,本文提出了一种新的可靠性度量标准 eMWTF。该标准全面考虑性能、面积、功耗开销对可靠性带来的负面影响。与传统的度量

标准相比,eMWTF 能够更加准确地定量表征微处理器的可靠性,因而更具指导意义,能够有效地指导设计探索和选择。

参考文献 (References)

- [1] Ziegler J F, Curtis H W, Muhlfield H P, et al. IBM experiments in soft fails in computer electronics (1978—1994) [J]. IBM Journal of Research and Development, 1996, 40(1): 3—18.
- [2] Mukherjee S. Architecture design for soft errors [M]. UK: Morgan Kaufmann Press, 2008.
- [3] Tang H H K. Nuclear physics of cosmic ray interaction with semiconductor materials: particle-induced soft errors from a physicist's perspective [J]. IBM Journal of Research and Development, 1996, 40(1): 91—108.
- [4] Freeman L B. Critical charge calculations for a bipolar SRAM array [J]. IBM Journal of Research and Development, 1996, 40(1): 119—129.
- [5] Truyen D, Boch J, Sagnes B, et al. Temperature effect on heavy-ion induced parasitic current on SRAM by device simulation: effect on SEU sensitivity [J]. IEEE Transactions on Nuclear Science, 2007, 54(4): 1025—1029.
- [6] 梁斌, 陈书明, 刘必慰. 温度对数字电路中单粒子瞬态脉冲的影响 [J]. 半导体学报, 2008, 29(7): 1407—1411. LIANG Bin, CHEN Shuming, LIU Biwei. Temperature dependence of digital single event transient [J]. Journal of Semiconductors, 2008, 29(7): 1407—1411. (in Chinese)
- [7] Mukherjee S S, Weaver C, Emer J, et al. A systematic methodology to compute the architectural vulnerability factors for a high-performance microprocessor [C]//Proceedings of IEEE/ACM International Symposium on Microarchitecture, 2003: 29—40.
- [8] Weaver C, Emer J, Mukherjee S S, et al. Techniques to reduce the soft error rate of a high-performance microprocessor [C]//Proceedings of International Symposium on Computer Architecture, 2004: 264—275.
- [9] Reis G A, Chang J, Vachharajani N, et al. Design and evaluation of hybrid fault-detection systems [C]//Proceedings of International Symposium on Computer Architecture, 2005: 148—159.
- [10] Gong R, Dai K, Wang Z Y. A framework to evaluate the trade-off among AVF, performance and area of soft error tolerant microprocessors [C]//Proceedings of IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems, 2008: 184—192.
- [11] Oh N, Shirvani P P, McCluskey E J. Control flow checking by software signatures [J]. IEEE Transactions on Reliability, 2002, 51(1): 111—122.
- [12] 龚锐, 陈微, 刘芳, 等. 一种软硬件结合的控制流检测与恢复方法 [J]. 计算机研究与发展, 2009, 46(2): 345—351. GONG Rui, CHEN Wei, LIU Fang, et al. Control flow checking and recovering by compiler signatures and hardware checking [J]. Journal of Computer Research and Development, 2009, 46(2): 345—351. (in Chinese)

使用位流重定位与差异配置在线演化数字系统*

姚睿,何坤,朱萍,李增武,羊宇中
(南京航空航天大学自动化学院,江苏南京 210016)

摘要:利用位流重定位与差异配置技术对现有基于动态部分重构的演化硬件实现方法进行改进,以解决其演化复杂电路时位流存储开销大和演化速度慢的问题。利用 Xilinx 早期获取部分重构技术,定制能实现位流重定位的可演化 IP 核。原始位流文件经设计形成算子核位流库存于外部 CF 卡上,方便系统调用。将现场可编程门阵列片内软核处理器 MicroBlaze 作为演化控制器,采用染色体差异配置技术,在线实时调节可演化 IP 核的电路结构,构成基于片上可编程系统的自演化系统。以图像滤波器的在线演化设计为例,在 Virtex-5 现场可编程门阵列开发板 ML507 上对系统结构和演化机制进行验证,结果表明,所提演化机制能有效节省位流存储空间,提高演化速度。

关键词:演化硬件;现场可编程门阵列;位流重定位;差异配置;自演化

中图分类号:TP302 **文献标志码:**A **文章编号:**1001-2486(2017)03-069-08

Online evolution of the digital system on bitstream relocation and discrepancy configuration

YAO Rui, HE Kun, ZHU Ping, LI Zengwu, YANG Yuzhong

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: To break through the limitations of huge memory space and low evolution speed for complex circuits' evolution, the bitstream relocation and the discrepancy configuration were adopted to improve the efficiency of the evolvable hardware implementation approach based on dynamic partial reconfiguration. Firstly, an evolvable IP core capable of bitstream relocation was customized by using the technology of early stage accession to partial reconfiguration provided by Xilinx. Then the original bitstream files were pre-synthesized to form a partial bitstreams library stored in the CF memory for the system to call. Next, a self-evolving system based on a programmable chip system was built, in which the soft processor, MicroBlaze, was utilized as the evolution controller. And the discrepancy configuration was adopted for the real-time adjustment of the circuit topology of the evolvable IP core. Finally, the system structure and the self-evolving mechanisms were verified by the online evolution of digital image filters implemented on the Xilinx Virtex-5 FPGA(field programmable gate array) development board ML507. Experimental results show that the proposed evolutionary mechanisms can reduce the storage space of bitstream files and can accelerate the speed of evolution significantly.

Key words: evolvable hardware; field programmable gate array; bitstream relocation; discrepancy configuration; self-evolution

在航空航天应用中,嵌入式系统必须承受恶劣的空间环境^[1],要求在提供复杂功能的同时,还需满足高可靠性、低功耗、低资源面积开销等需求。这些要求互相制约,使系统复杂度以指数增长,为系统设计带来极大挑战。迫切需要能随环境和需求变化自动调节自身结构和行为以实现任务目标的自适应系统^[2-3]。然而,传统自适应系统的适应能力有限(如算法结构固定、仅参数可变等),且无法实现故障情况下的自主修复,因而已无法满足现代自适应系统的需求。演化硬件(Evolvable HardWare, EHW)的出现为构建自适应系统提供了一种解决方案。以演化算法(Evolutionary Algorithm, EA)为全局搜索的主要工具,以现场可重构器件为评估平台和实现载体,寻求在不依赖先验知识和人工干预的情况下,通过演化来获得满足给定要求的电路和系统结构^[4-5],进而根据环境变化自主调节自身结构及功能,达到从故障中恢复、在运行生命周期内提高性能等目的。

早期的 EHW 通过软件仿真离线演化,仅将最终所得最优染色体配置于可编程硬件器件上进行验证,称作外部演化^[6]。该方法比较适合研究演化方法,探索新型 EHW 结构模型,但不能实时调整硬件电路结构,无法满足系统自适应需求。

* 收稿日期:2016-01-12
基金项目:国家自然科学基金资助项目(61402226);中央高校基本科研业务费专项资金资助项目(NS2014036)
作者简介:姚睿(1974—),女,河南邓州人,副教授,博士,硕士生导师,E-mail: yaorui@nuaa.edu.cn

随着技术的发展,出现了内部演化方式,直接将每代种群的每条染色体分别下载到可重构器件中进行评估,因而可实时调整硬件结构,为实现自适应硬件提供了可能。

20 世纪 90 年代中期,EHW 的思想第一次在 Xilinx 的 XC6200 系列现场可编程门阵列(Field Programmable Gate Array, FPGA)上实现。该芯片内部结构位串(位流)格式公开,可直接对流操作,很适合实现演化,但已于 1998 年停产。此后,由于担心随机修改位流威胁器件的完整性,商用 FPGA 芯片位流格式不再公开;厂商提供的重构技术也不够成熟,故无法在商用 FPGA 上直接操作位流进行无约束演化。为此,Seikanina 提出了基于虚拟可重构电路(Virtual Reconfigurable Circuit, VRC)的 EHW 实现方式^[7-9]。该方法在 FPGA 上实现由处理节点矩阵组成的虚拟可重构层,每个节点包含所有需求功能,可通过多路选择器选择;同时,利用 FPGA 片上微处理器核运行 EA,实现了基于片上可编程系统(System On a Programmable Chip, SOPC)的自演化系统。该方法能实现电路结构与功能的自调整,以适应外部环境变化和故障的自恢复^[10-12],成为实现自适应系统的一种解决方案。然而,由于 VRC 中每个节点同时实现了该节点所有可能实现的功能,资源开销较大,且多路选择器加大了电路延时^[13-14]。

因此,基于动态部分重构(Dynamic Partial Reconfiguration, DPR)的方法应运而生。该方法利用 Xilinx 早期获取部分重构技术,在 FPGA 上使用规则的二维可重构分区(Reconfigurable Partition, RP)阵列取代 VRC 定制可演化 IP 核^[15]。每个 RP 可配置为多种功能,每种功能可用一个函数描述。设计阶段预先产生包含该函数信息的位流文件,称作算子核位流;所有位流存放于同一存储空间,形成算子核位流库。演化过程中,RP 间连线固定,由 EA 控制改变各 RP 配置的算子核位流,以实现不同电路结构。与 VRC 方法相比,该方法未使用多路选择器,减少了电路延时;也未同时在每个单元上实现所有可能的功能,每个 RP 所占用芯片面积取决于其中最复杂的功能,减少了面积开销。然而,文献[15]中建立算子核位流库时,要为每个 RP 的每种功能生成位流并进行存储,因此存储空间开销大;且评估每一条染色体的适应度时,无论各 RP 模块功能改变与否,均对其重新配置,演化速度较慢。

因此,本文采用位流重定位和差异配置技术改进现有基于 DPR 的 EHW 实现方式,以降低位

流存储量和提高演化速度。

1 系统的 SOPC 体系结构

1.1 总体结构与工作原理

系统的 SOPC 体系结构如图 1 所示,包含一个片上软核处理器 MicroBlaze,所有 IP 核以外设形式挂接于处理器本地总线(Processor Local Bus, PLB)上。MicroBlaze 作为演化控制器,运行 EA 控制演化过程;自定制可演化核 MATH_IP_Core 可根据演化命令调节内部处理功能,实现系统功能的自适应;Input_data 和 Idealout_data 是定制的 ROM IP 核,分别存放演化区域的输入数据和期望输出数据;演化过程所需算子核位流与系统全局初始化位流一起存放于外部 CF 卡上;系统高级配置环境(Advanced Configuration Environment, ACE)控制器负责从 CF 卡读取位流文件;重构引擎硬件内部配置访问端口(HardWare Internal Configuration Access Port, HWICAP)负责 FPGA 重构的实现;通用异步收发传输器(Universal Asynchronous Receiver/Transmitter, UART)用于实现 FPGA 开发板与 PC 机超级终端的通信。

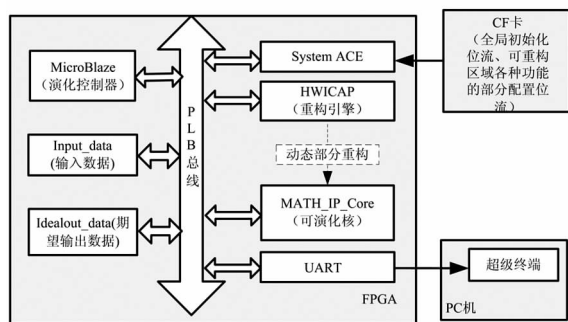


图 1 自演化数字系统的总体结构框图

Fig. 1 General structure of the self-evolving system

开发板上电后, System ACE 通过联合测试工作组(Joint Test Action Group, JTAG)接口读取 CF 卡上的 ace 文件,实现系统的全局初始化。演化开始时, MicroBlaze 根据染色体配置向 HWICAP 发送重构命令; HWICAP 通过 System ACE 从 CF 卡读取相应算子核位流,配置到对应 RP 中,实现染色体到 FPGA 底层硬件的映射;然后将输入只读存储器(Read Only Memory, ROM)核中的数据作为 MATH_IP_Core 的输入,并将运算结果反馈给 EA,与期望输出数据比较,评估个体适应度;接着 EA 产生下一代种群;重复以上操作,即可实现系统的自演化。演化结果可通过 PC 机超级终端观察。

1.2 可演化核结构

只要硬件资源允许,可演化核可以设计为任意 $m \times n$ 的 RP 阵列,每个 RP 可根据需要配置为任意多种功能,如图 2 所示。每个 RP 模块有两个输入端口和两个完全相同的输出端口,阵列中每个 RP 可配置任意 p 种功能。与 VRC 方法相比,该方法未使用大量的多路选择器,也未在每个节点同时实现所有可能的功能,各 RP 占用芯片面积取决于其中最复杂的功能,减少了电路延时和面积开销。此外,该结构也可根据目标电路的复杂度方便地扩展,提高了系统设计的灵活性。

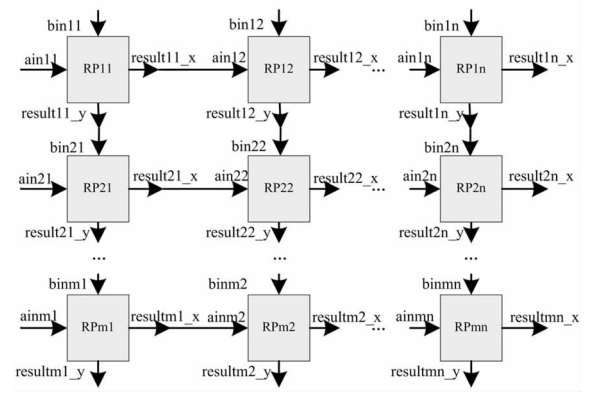


图 2 可演化核的二维阵列结构

Fig. 2 2D structure of the evolvable core

目前基于 DPR 思想演化时,需要预先为每个 RP 的每种功能生成位流文件。若每个 RP 可实现 p 种功能,则需为其生成 p 个位流文件;这样 $m \times n$ 的 RP 阵列需要 $m \times n \times p$ 个位流,大大增加了位流存储空间。因此,本文仅针对某一 RP 生成实现 p 种功能所需的 p 个位流文件并存储;演化过程中,通过位流重定位技术,实现其他 RP 的功能,降低了对存储空间的要求。

另外,目前基于 DPR 技术演化过程中,无论每个 RP 的功能是否改变,均对演化区域中所有 RP 的位流进行重新配置,大大增加了演化时间,降低了演化速度。实际上,每次演化需要改变的 RP 的个数有限,尤其在演化后期,仅很少一部分 RP 的功能需要改变。因此,没有必要每次对所有的 RP 进行完全配置。为此,本文提出了差异配置技术,配置过程中通过对比新染色体与原染色体的差异,仅重新配置需要改变的 RP,大大减少了配置时间,提高了演化速度。

2 位流重定位技术

2.1 Virtex-5 FPGA 位流格式

位流文件(bit 文件)是 FPGA 的配置数据流,

其中包含了配置命令字和配置数据。它是一个二进制文件,包括文件头和 FPGA 的有效配置数据^[21],其构成如图 3 所示。文件头主要表示 FPGA 的类型及文件生成的时间信息。尽管不同位流文件的文件头长短和内容不同,但是有效数据总是以同步字“AA995566”开始。



图 3 bit 文件的组成示意图

Fig. 3 Diagram of the bit file

位流文件中有效数据可分为 3 个功能区:配置命令字区、配置数据区和循环冗余校验(Cyclic Redundancy Check, CRC)区。配置命令字区的作用是利用内部配置逻辑来加载数据帧;配置数据区是实现功能的配置数据帧;CRC 校验区的作用是完成初始化启动及 CRC 校验。只有 CRC 校验完成之后,位流才能配置到 FPGA 中。

2.2 位流重定位的原理

利用传统 DPR 技术设计可重构系统时,即使实现同一种功能,每个 RP 的部分位流均不同,故一个 RP 的位流不能直接配置到其他 RP 上。图 4 所示系统包含 3 个 RP,每个 RP 均可实现加法(adder)和减法(sub)两种功能。若采用传统 DPR 技术,PRR1 的位流文件 adder_1 只能用于配置 RP1,而不能配置 RP2 和 RP3。因此,每个 RP 需要 2 个位流文件,一共需要 6 个位流文件,存储空间的开销较大。

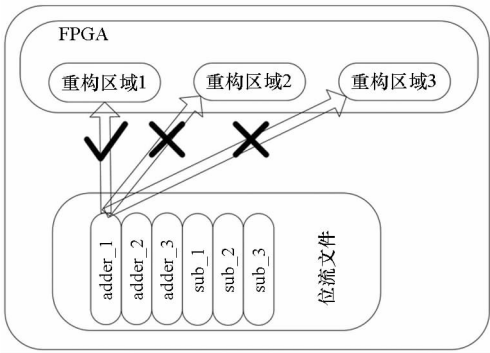


图 4 DPR 系统位流配置示意图

Fig. 4 Diagram of the DPR system's bitstream configuration

位流重定位是指在传统 DPR 设计流程的基础上,通过增加 RP(可重构区域)设计约束,并对位流进行适当修改,允许一个部分位流(部分重构模块)从一个 RP 配置到资源、区域大小相同的另外一个 RP。

RP 满足设计约束时,比较不同位置 RP 上实现相同功能的位流发现,其配置数据区相同,只有帧地址寄存器(Frame Address Register ,FAR)和 CRC 值不同。因此,实现位流重定位时,只需修改配置位流中的 FAR 和 CRC 的数据。如若需将图 4 中 RP1 的位流文件 adder_1 配置到 RP2,首先应要求 RP2 和 RP1 满足设计约束,其次需要根据 RP2 与 RP1 的相对位置信息修改 adder_1 中 FAR 和 CRC 的值。

2.3 位流重定位的实现

若各 RP 满足设计约束,要将某 RP 的某种功能的位流文件重新定位到其他 RP 模块,只需修改 FAR 和 CRC 的值,配置数据区无须改变。

2.3.1 FAR 结构

FAR 用于存储位流配置的起始地址,其有效数据包括 5 部分:配置块类型、上半部/下半部、行地址、列地址、次地址,如图 5 所示。Virtex - 5 FPGA 包含可编程输入/输出块,可配置逻辑块,块 RAM、CLK、DSP 等资源,分别对应不同编码(“001”代表块 RAM 互连,“000”代表其他资源)。资源对称分成上下两部分,行编号分别从中间向顶部和底部递增,列编号从左向右递增,次地址为每列包含的地址。

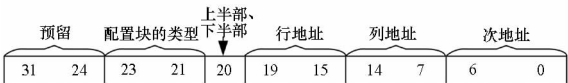


图 5 FAR 寄存器的结构
Fig. 5 Structure of the FAR

2.3.2 CRC 值的计算

CRC 是为了检查配置文件的合法性,进而保护 FPGA 设备的安全而对配置数据进行的循环冗余校验。虽然重定位位流中配置数据区内容未改变,但是 FAR 的值发生了改变,因此必须重新计算其 CRC 的值。Virtex - 5 FPGA 配置位流的 CRC 值可根据式(1)计算。

$$x^{32} + x^{28} + x^{27} + x^{26} + x^{25} + x^{23} + x^{22} + x^{20} + x^{19} + x^{18} + x^{14} + x^{13} + x^{11} + x^{10} + x^9 + x^8 + x^6 + 1 \quad (1)$$

CRC 值的计算可以在线或离线进行。为了减少重构时间,本文采用离线计算的方式,根据式(1)计算重定位位流中 FAR 值改变后对应的 CRC 值并存储。演化过程中实现位流重定位时不用在线计算 CRC 值,从而提高演化速度。

3 自演化系统的设计实例与实现结果

本节以在 Virtex - 5 FPGA 开发板 ML507 上

实现图像滤波器的在线演化设计为例,说明系统的软、硬件设计方法与实现结果。

3.1 自演化系统的硬件设计

3.1.1 系统硬件平台的设计

首先在赛灵思平台工作室(Xilinx Platform Studio, XPS)中创建系统硬件平台,然后将生成的 system. ngc 文件导入 PlanAhead 中,实现 DRP 设计。

3.1.2 可演化 IP 核的设计

设计图像滤波器在线演化系统时,可演化核为 3×3 二维阵列结构。每个 RP 有两个 8 bit 输入和两个完全相同的 8 bit 输出,可配置为表 1 所示的 8 种功能。PR 间连线固定,改变各 RP 的配置可实现不同电路结构和系统功能。采用 3×3 窗口采集图像数据,作为 RP 阵列的输入;最后一个 RP 的输出作为最终运算结果。

表 1 功能单元及其对应编码
Tab. 1 Functional units and their encoding

名称	功能	对应编码		
		bit[3i+2]	bit[3i+1]	bit[3i]
adder	ain + bin	0	0	0
absolute	ain - bin	0	0	1
aiden	ain	0	1	0
ainver	255 - ain	0	1	1
binver	255 - bin	1	0	0
average	(ain + bin)	1	0	1
max	max(ain, bin)	1	1	0
min	min(ain, bin)	1	1	1

注:i=0,1,⋯,8;n=i+1。

3.1.3 可重定位位流设计

位流重定位技术对各 RP 区域的设计,包括各 RP 模块区域划分、各 RP 与静态区域接口以及布线路径等,均有特殊要求。要求重构区域设计必须满足以下规则:①重构区域所占的逻辑资源相同;②重构模块端口引脚的数目一致;③重构模块代理逻辑的相对位置一致;④静态区域的布线不能穿过动态区域。这些规则均可在 PlanAhead 中通过修改约束文件实现。

设计过程主要包括:规划 RP 区域、约束接口代理逻辑位置和统一接口布线信息。

规划重构区域时,必须保证每个 RP 的高度为整行,最小 RP 为一行一列。若 RP 高度小于 1 行,则位流将会包含静态区域的配置信息。一个 RP 所需资源和资源利用率如图 6 所示。

Physical Resource Estimates			
Site Type	Available	Required	% Util
LUT	160	0	0
FD_LD	160	14	9
SLICEL	40	5	13

图6 1个RP资源利用情况

Fig. 6 Resource usage of one RP

可重构区域与静态区域的接口采用代理逻辑(proxy logic)实现。代理逻辑可通过实现工具自动插入。要实现位流重定位,各RP区域代理逻辑的相对位置必须一致,故需修改约束文件来修改代理逻辑的位置。

代理逻辑放置于可重构区域,它与静态区域之间的布线路径信息也包含在重构位流信息内。为保证所有RP的代理逻辑与静态区域之间的布线路径信息一致,还需对所有RP的代理逻辑和静态模块之间的布线信号路径进行分析和修改。首先提取所有RP模块的路径信息,然后选取其中一个RP的信号路径信息作为标准,对其他RP区域的路径信息进行修改。

3.1.4 图像数据的存储

演化图像滤波器时,需要存储待滤波图像数据、理想图像数据和滤波后图像数据。为方便管理,本文将待滤波和理想图像数据存储于FPGA片内块随机存取存储器(Block Random Access Memory, BRAM)中,并将其定制为ROM IP核,挂接于PLB总线上,作为外设调用;滤波后的图像数据作为变量,临时分配存储空间。演化过程中,将待滤波图像数据作为可演化核的输入。读取滤波后图像数据,按照一定的准则与理想图像数据比较,即可得到适应度值。

3.2 自演化系统的软件设计

系统软件设计的主要任务是在MicroBlaze上运行演化算法实现演化过程的控制。本文选择遗传算法进行系统软件设计。

3.2.1 染色体编码方式

为降低演化复杂度,本文采用加、减、求均值等函数功能模块作为演化积木块,采用间接编码方式。3×3 RP阵列的分段二进制编码方案如图7所示,RP1~RP9分别对应9个RP的编码。每个RP的8种功能需3位二进制数描述,故与RP阵列功能相关的染色体长度为3×9=27 bit。各RP编码与功能对应关系如表1所示。3×3窗口每次需要9个像素值,故每个输入端口的选择需4位二进制数描述。由于3×3 RP阵列共有6个输入端口,

故与输入选择相关的染色体长度为4×6=24 bit。因此染色体总长度为27+24=51 bit。

RP[]	RP9		RP8		RP7		RP6		RP5		RP4		RP3		RP2		RP1	
	bit[]		bit[]		bit[]		bit[]		bit[]		bit[]		bit[]		bit[]		bit[]	
	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9
	8	7	6	5	4	3	2	1	0									

图7 3×3 RP阵列功能的染色体编码方案

Fig. 7 Chromosome encoding of the 3×3 RP array

3.2.2 染色体差异配置技术

评价染色体时,需将其解码后配置到可演化核的重构区域中。目前基于DPR的方法采用整体配置,即对每个RP功能和所有输入端口均进行重新配置,增加了配置时间。尤其是改变各RP功能时,从外部读取位流和通过内部配置访问端口(Internal Configuration Access Port, ICAP)端口配置耗时较多。因此,为了减少配置时间,提高演化速度,本文提出了染色体差异配置技术。配置新染色体时,调用相应算子位流之前,首先逐位区比较待配置染色体与当前染色体,找出不同的位区,并重新配置对应的RP区域,而功能不变的RP区域则不进行配置。

实现过程中,首先利用配置矩阵描述新、老染色体,然后对各矩阵元素进行异或,得到差异矩阵,最后根据差异矩阵和染色体编码方法对需改变配置的RP区域进行重构,实现差异配置。

染色体差异配置技术的采用大大减少了需要配置区域的数目,进而减少了演化的时间。

3.2.3 适应度计算

可演化核有6个输入1个输出。滤波过程与典型的图像卷积滤波器类似,采用3×3窗口选取每个像素及其相邻像素作为可演化核的输入,可演化核输出即为中心像素滤波后的输出。本文采用的滤波图像像素为125列124行,除了第1行、第124行、第1列及第125列等“图像边框”中像素外,其他像素点均可作为中心点。衡量滤波效果时,采用式(2)所示的平均每像素误差(Mean Difference Per Pixel, MDPP)作为评价标准,

$$MDPP = \frac{\sum_{i=0}^{R-1} \sum_{j=0}^{C-1} |ideal(i,j) - filtered(i,j)|}{C \times R}$$

(2)

式中,C为列数,R为行数,ideal(i,j)为理想无噪图像像素,filtered(i,j)为滤波后的像素。MDPP越小表明滤波图像效果越好。计算每一代种群中每个个体对应的MDPP值,并保留MDPP值最小的个体,参与下一代种群的产生。

3.2.4 遗传算子设计

遗传算法通过选择、交叉和变异等遗传算子

产生新个体。本文为了降低复杂度,对算法进行了简化,仅采用了选择算子和变异算子。

1) 选择算子

选择算子采用联赛选择方式。每次联赛选择时,首先从父代种群中随机选取一定数量的个体,选择其中适应度最大者进入下一代种群;进行 P (种群规模) 次选择即可产生新的种群。该过程中,适应度较高的个体可能多次被选中,而适应度较低的个体可能一次未被选中,充分体现了自然界生物进化过程中“优胜劣汰”的原则。

2) 变异算子

变异算子以一定的概率对染色体位串进行翻转。设变异概率为 P_M ,则种群中任意个体被选中参与变异操作的概率为 P_M 。本文设计采用简单均匀随机变异操作。为保证个体变异后与其父代的差异不会过大,变异率一般取值较小,为 0.1 或更小,以保证种群发展的稳定性。

3.2.5 自演化操作过程

自演化操作流程如图 8 所示,有两个演化终止条件:一是找到达到最大适应度的个体,即找到最优解;二是达到设定的最大演化代数。

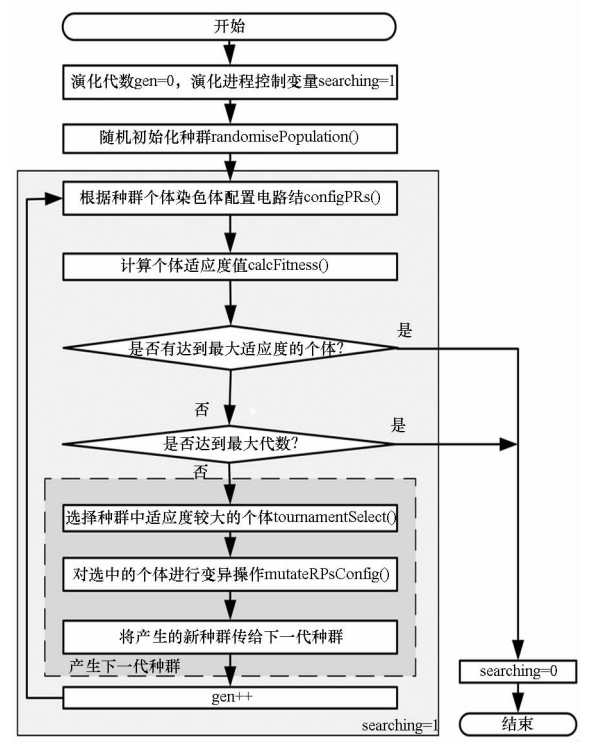


图 8 自演化操作流程图
Fig.8 Operation flow of self-evolution

4 实验结果与分析

4.1 位流重定位技术对位流存储空间的影响

采用位流重定位技术可以减少需要存储的位

流个数,节省存储空间。表 2 所示为实现图像滤波器在线演化时,采用位流重定位技术(本文方法)与不采用位流重定位技术(传统方法)所需位流个数和存储空间大小的对比。由表 2 可见,本文方法所需位流文件数目比传统方法减少了 64 个。每个位流文件大小为 7 KB,故与传统非重定位方法相比,采用位流重定位技术可以节约 448 KB 的存储空间,比传统方法节省 88.9%。

表 2 所需位流文件的个数和存储空间对比
Tab.2 Comparisons of number of bit files and memory space

逻辑功能	位流文件个数		所需存储空间大小	
	本文方法	传统方法	本文方法	传统方法
adder	1	9	7 KB	63 KB
absolute	1	9	7 KB	63 KB
average	1	9	7 KB	63 KB
aiden	1	9	7 KB	63 KB
max	1	9	7 KB	63 KB
min	1	9	7 KB	63 KB
binver	1	9	7 KB	63 KB
ainver	1	9	7 KB	63 KB
总计	8	72	56 KB	504 KB

事实上,采用位流重定位技术所节约存储空间的大小取决于 RP 的个数及 RP 区域的大小。本文 3×3 RP 阵列中共有 9 个 RP,采用位流重定位技术仅需为一个 RP 生成 8 种逻辑功能的位流文件;而非重定位方法则需为 9 个 RP 各自生成 8 种位流文件(共 $8 \times 9 = 72$ 个),故存储空间节约率为 $8/9$ 。若采用 $m \times n$ 的 RP 阵列,每个 RP 可实现 p 种功能,则非重定位方法需存储 $m \times n \times p$ 个位流,假设每个位流为 s KB,则需 $m \times n \times p \times s$ KB 存储空间;而位流重定位技术仅需存储 p 个位流,即 $p \times s$ KB 的存储空间,为传统方法的 $1/(m \times n)$,即存储空间节约率为 $(m \times n - 1)/(m \times n)$ 。

4.2 差异配置技术对演化时间的影响

种群规模 $P = 64$, 联赛规模 $T = 10$, 变异率 $P_M = 2/256$ 时,采用本文的差异配置与完全配置的演化时间与演化代数关系对比如图 9 所示。由图 9 可见,完全配置方法的演化时间与演化代数成正比;原因是该方法对每代的每条染色体均进行完全配置,故配置所需时间基本固定。

而差异配置在评估时将新染色体与原染色体进行比较,仅重新配置功能改变了的 RP,故大大节省了每条染色体的配置时间,提高了演化速度。而且,演化的代数越多,差异配置的优势越明确。

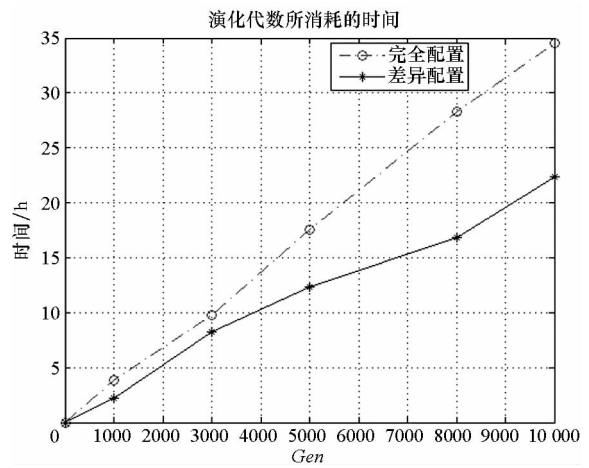


图9 差异配置和完全配置演化时间对比

Fig.9 Comparison of consumed time between discrepancy configuration and complete configuration

4.3 图像滤波的效果

采用 125 × 124 具有 $L = 256$ 级灰度的标准“Lena”图像作为测试图像,分别注入均值为 0,方差为 0.03 的高斯噪声和噪声强度为 5% 的椒盐噪声,进行图像滤波器在线演化设计,滤波前后图像如图 10 所示,滤波前后 MDPP 值对比如表 3 所示。实验中遗传算法的参数:种群规模 $P = 64$,联赛规模 $T = 10$,变异率 $P_M = 2/256$,演化终止代数 $Gen = 10\ 000$ 。

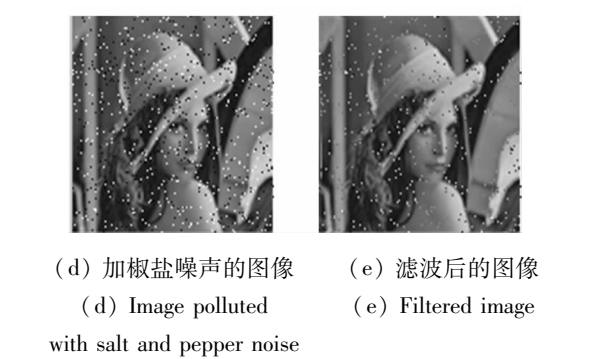
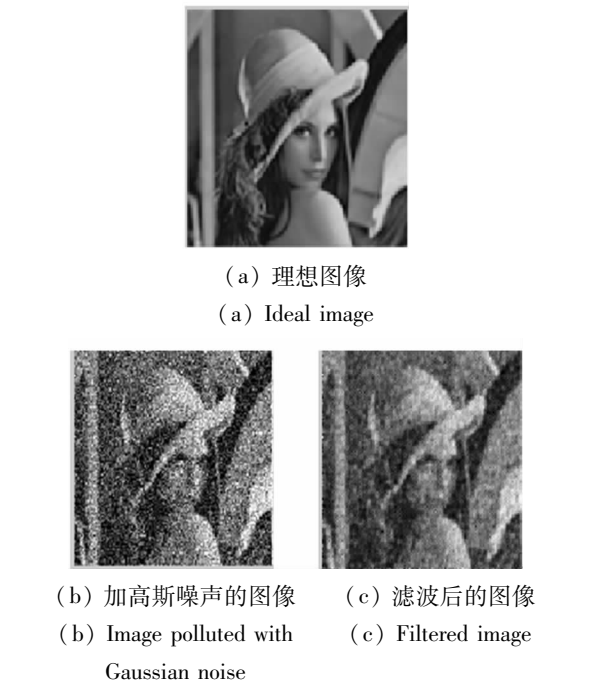


图 10 滤除图像中高斯噪声的结果

Fig. 10 Experimental results of filters Gaussian noise

表 3 滤波前后 MDPP 值对比		
Tab. 3 Comparison of MDPP before and after filtering		
噪声	滤波前 MDPP	滤波后 MDPP
均值 0、方差 0.03 的高斯噪声	33.08	14.86
5% 的椒盐噪声	6.55	1.21

由图 10 和表 3 可见,本文系统可以对高斯噪声和椒盐噪声进行有效滤波。尤其是对椒盐噪声滤波效果更为明显。滤波前图像的 MDPP 值为 6.40,滤波后仅为 1.21,有效降低了 MDPP 值。

5 结论

为克服现有演化硬件实现方式存在的不足,本文采用位流重定位和差异性配置技术实现自演化数字系统,研究了其体系结构与在线自主演化技术。与目前基于 VRC 的 EHW 实现方式相比,本文方法未利用大量的多路选择器,也未同时在每个可重构模块上实现所有可能的功能,减少了电路延时,降低了器件资源面积开销。与现有的基于 DPR 的实现方式相比,本文方法大大减少了存储位流文件所需要的存储空间,并大大缩短了演化每代所需要的时间,提高了演化速度。

参考文献 (References)

[1] Vipin K, Fahmy S A. Mapping adaptive hardware systems with partial reconfiguration using CoPR for Zynq [C]// Proceedings of Adaptive Hardware and Systems (AHS), NASA/ESA Conference on Adaptive Hardware and Systems (AHS - 2015), 2015: 1 - 8.

[2] Mora J, Gallego A, Otero A, et al. A noise-agnostic self-adaptive image processing application based on evolvable hardware [C]// Proceedings of Design and Architectures for Signal and Image Processing (DASIP), 2013: 351 - 352.

[3] Gallego A, Mora J, Otero A, et al. A self-adaptive image processing application based on evolvable and scalable hardware [C]// Proceedings of Field Programmable Logic and Applications (FPL), 2013.

- [4] 莫宏伟, 徐立芳. 基于 Memetic 算法的电路演化设计研究[J]. 电子学报, 2013, 41(5): 1036 – 1040.
MO Hongwei, XU Lifang. Research on evolvable hardware design based on Memetic algorithm [J]. Acta Electronica Sinica, 2013, 41(5): 1036 – 1040. (in Chinese)
- [5] 刘少腾, 来金梅, 陈利光, 等. 可进化可重构图像滤波器的设计[J]. 复旦学报: 自然科学版, 2010, 49(6): 709 – 715.
LIU Shaoteng, LAI Jinmei, CHEN Liguang, et al. Design of an evolvable reconfigurable image filter[J]. Journal of Fudan University: Natural Science, 2010, 49(6): 709 – 715. (in Chinese)
- [6] Haddow P C, Tyrrell A M. Challenges of evolvable hardware: past, present and the path to a promising future[J]. Genetic Programming and Evolvable Machines, 2011, 12(3): 183 – 215.
- [7] Sekanina L. Virtual reconfigurable circuits for real-world applications of evolvable hardware[M]//Tyrrell A M, Haddow P C, Torresen J. Evolvable Systems: From Biology to Hardware, Germany: Springer Berlin Heidelberg, 2003: 186 – 197.
- [8] Sekanina L. Evolutionary design of gate-level polymorphic digital circuits[M]//Rothlauf F, Branke J, Cagnoni S, et al. Applications of Evolutionary Computing, Germany: Springer Berlin Heidelberg, 2005: 185 – 194.
- [9] Sekanina L, Harding S L, Banzhaf W, et al. Image processing and CGP [M]. Cartesian Genetic Programming, Germany: Springer Berlin Heidelberg, 2011: 181 – 215.
- [10] 朱继祥, 李元香, 邢建国. 可重构系统的演化修复机制[J]. 计算机学报, 2014, 37(7): 1599 – 1606.
ZHU Jixiang, LI Yuanxiang, XING Jianguo. The evolvable recovery of reconfigurable system [J]. Chinese Journal of Computers, 2014, 37(7): 1599 – 1606. (in Chinese)
- [11] Zhang X, Luo W. Evolutionary repair for evolutionary design of combinational logic circuits [C]//Proceedings of IEEE Congress on Evolutionary Computation (CEC), 2012: 1 – 8.
- [12] 姚睿, 王友仁, 于盛林, 等. 具有在线修复能力的强容错三模冗余系统设计及实验研究[J]. 电子学报, 2010, 38(1): 177 – 183.
YAO Rui, WANG Youren, YU Shenglin, et al. Design and experiments of enhanced fault-tolerant triple-module redundancy systems capable of online self-repairing [J]. Acta Electronica Sinica, 2010, 38(1): 177 – 183. (in Chinese)
- [13] Salvador R, Otero A, Mora J, et al. Self-reconfigurable evolvable hardware system for adaptive image processing[J]. IEEE Transactions on Computers, 2013, 62(8): 1481 – 1493.
- [14] Salvador R, Otero A, Mora J, et al. Implementation techniques for evolvable HW systems: virtual vs dynamic reconfiguration [C]//Proceedings of International Conference on Field Programmable Logic and Applications (FPL), 2012: 547 – 550.
- [15] 孙艳梅. 基于 FPGA 动态部分重构的数字在线演化技术研究[D]. 南京: 南京航空航天大学, 2015.
SUN Yanmei. Research on online evolution technology of digital system based on dynamic partial reconfiguration of FPGA [D]. Nanjing: Nanjing University of Aeronautics and Astronautics, 2015. (in Chinese)

日常交互中朋友关系强度度量方法*

史殿习,杨若松,莫晓赞,李 寒,赵邦辉
(国防科技大学 计算机学院,湖南 长沙 410073)

摘 要:针对如何度量日常生活中人们之间的关系强度问题展开研究,提出一个从日常轨迹、语义位置以及语义标签三个层次度量朋友之间关系强度的层级模型 FRSHV。采用动态时间规整模型通过计算朋友之间的空间距离来度量其日常轨迹之间的相似度,进而使用轨迹序列熵值对用户每天轨迹的相似度进行加权处理,将其作为朋友之间的关系强度;采用主题模型隐含狄利克雷分布分别计算朋友之间的基于语义位置和语义标签的行为模式的相似性,将其作为朋友之间的关系强度;采用集成学习的思想对三个层次的度量结果进行投票,以投票结果作为最终的朋友之间的关系强度。在公开数据集上对 FRSHV 模型的有效性进行实验验证,结果表明该模型能够有效地度量朋友之间的关系强度。

关键词:关系强度;轨迹相似度;动态时间校正;熵;潜狄利克雷分布;投票
中图分类号:TP391 **文献标志码:**A **文章编号:**1001-2486(2017)03-077-08

Measuring method for friend relationship strength in daily communication

SHI Dianxi, YANG Ruosong, MO Xiaoyun, LI Han, ZHAO Banghui
(College of Computer, National University of Defense Technology, Changsha 410073, China)

Abstract: The FRSHV (friend relationship strength hierarchy vote), a hierarchical model, was proposed to measure the friend relationship strength by user's daily moving track, semantic positions and the corresponding semantic labels. The daily track similarity was measured by dynamic time warping model using the spatial distance between friends, and the results were then weighed by the entropy of track series. The similarities of friend's behavior patterns were inferred by latent Dirichlet allocation topic model, respectively using semantic positions and the corresponding semantic labels. Finally, these three similarity results were voted for the ultimate relationship strength. The FRSHV was evaluated by using an open dataset and the results proved the validity of the model in inferring friend's relationship strength.

Key words: relationship strength; trajectory similarity; dynamic time warping; entropy; latent Dirichlet allocation; vote

目前,内嵌了各种各样传感器的智能手机已经成为人们日常生活中集通信、计算及感知于一体的移动平台,通过内嵌的各种传感器如 GPS、加速度、麦克风等可以随时随地感知和获取人们自身及其周围环境的各种信息。通过智能手机所收集的各种数据研究人们之间的日常交互行为和人们之间的社会关系成为普适计算领域当中一个重点研究的问题。文献[1]基于手机所收集的各种数据推理人们之间的社会交互关系以及群组的活动韵律,从而洞察个人和组织的行为模式;文献[2]研究分析了家庭和朋友圈对个体行为在社交方面的影响;文献[3]研究了在校学生的日常活动、交互情况、精神健康与学业成绩之间的关系;文献[4]则从多渠道、细粒度地收集反映在校

学生日常活动和交互情况的各种数据,从多个层面真实、全面地反映学生日常活动以及他们之间的交互行为和交互关系。但是,这些研究重点关注的是人们之间的日常交互行为和交互关系。关系强度度量的是人们之间的亲密程度,通过关系强度可以更好地了解人们之间关系的强弱,进而了解人们之间的亲密程度,从而可以更好地预测社会关系的演变以及社交结构的变化、促进信息传播以及传染疾病的预防与控制等。

社会关系强度理论始于文献[5]中对于弱关系的研究,将弱关系和强关系的测量分为四个维度,即交往人员之间的互动频率、感情的投入程度、关系亲密程度和在互惠互利上的交换程度;文献[6]对这四个维度做了相关指标化;文献[7]认

* 收稿日期:2016-02-10
基金项目:国家自然科学基金资助项目(61202117,91118008)
作者简介:史殿习(1966—),男,山东龙口人,教授,博士,博士生导师,E-mail:dxshi@nudt.edu.cn

为关系强度涉及关系的数量以及交往的频率。随着关系强度研究领域的不断发展,以互动频率、联系次数、亲密程度为关系强度核心测量指标的主流研究观点^[8]逐渐形成。但是,如何度量社会网络中人们之间的关系强度一直是社交网络关系分析中的一个难点问题。

通过智能手机可以随时随地地获取位置、通话记录、短信、微信等体现人们之间日常交互和社会关系的各种信息,而人们之间的交互频率、时间、位置、地点、距离以及轨迹相似性等信息能够直接体现人们之间的交互关系以及关系强度,因为关系密切的人们之间更愿意面对面地进行交流,如朋友之间就会经常进行面对面的交流(如聚会、一起游览等)。通过对这些信息的分析处理,可以更好地度量朋友之间的关系强度。为了方便描述,本文将分析处理的对象称为用户,并认为用户和陌生人之间的关系强度应该为零如果他们互不认识。虽然对一个用户来说,即使其与一些陌生人并不认识,其也可能会经常与之在一些地方同时出现,但本文只考虑用户和其好友之间的关系强度。本文设想能够在一定程度上反映两个朋友之间的关系,而非完整全面地度量两个用户之间的关系;认为使用手机上所有传感器的全部数据能够精确地分析朋友之间的关系强度。轨迹数据是手机传感器数据非常重要的组成部分,本文主要研究如何只使用轨迹数据度量朋友之间的亲密程度。文献[9]认为用户之间的关系强度与用户共同出现的时间和共同出现的位置相关,提出了一个基于 GPS 轨迹数据的层级模型,根据用户的 GPS 轨迹来度量用户之间的关系强度,并在仿真数据集上进行了实验验证。

本文在文献[9]的基础上,针对如何度量日常生活中人们之间的关系强度问题展开研究,提出了一个可以对 GPS 数据和基站数据进行处理,从日常轨迹、语义位置以及语义标签三个层次度量用户与朋友之间关系强度的层级(Friend Relationship Strength Hierarchy Vote, FRSHV)模型。该模型采用动态时间校正(Dynamic Time Warping, DTW)模型通过计算用户与朋友之间的空间距离来度量其轨迹之间的相似度,进而使用轨迹序列熵值对用户每天轨迹的相似度进行加权处理,并将其作为用户与其朋友之间的关系强度;采用主题模型潜狄利克雷分布(Latent Dirichlet Allocation, LDA)分别计算用户与朋友之间的基于语义位置和语义标签的行为模式的相似性,将其作为用户与朋友之间的关系强度;采用集成学

习的思想对三个层次的度量结果进行投票,以投票结果作为最终的用户与朋友之间的关系强度。

1 关系强度度量方法

通过对社会心理学相关研究成果的分析,认为人们之间的关系强度与他们之间的轨迹相似性以及日常行为的相似性密切相关,因此,为了有效地度量人们之间的关系强度,本研究从人们之间的日常轨迹和日常行为这两个角度出发,提出采用不同计算方法来计算人们之间的关系强度。

1.1 基于 DTW 模型的计算方法

空间距离能够直观反映人们之间在物理世界中的距离,空间距离非常接近的用户在现实生活中会有更多的面对面的交互,从而增强两个人之间的关系强度。根据社会心理学的研究成果,文献[10]在一个大型住宅区研究了接近性效应(接近性效应指两个人住得越近越可能是朋友),结果表明人们居住得越近,不管这种近是物理距离还是功能性距离,人们越容易称为朋友。文献[11]用实验证实了单纯接触效应,即熟悉性能够促进好感,实验结果表明接触频率越高喜欢程度越强。

DTW 是 Itakura 等于 1987 年^[12]提出的一种距离度量方法。将用户的轨迹数据看作一个时间序列,因此同样可以使用 DTW 方法度量轨迹的相似度,并且将轨迹相似度作为人们之间的关系强度。通过深入分析 DTW 算法可知,序列的长度越长,则距离可能越大。因此,采用文献[13]中的三种归一化方法对 DTW 的计算结果进行进一步的处理和优化,即采用 DTW 结果除以最优变形路径的长度、DTW 结果除以两个序列中较短序列的长度以及 DTW 结果除以两个序列中较长序列的长度三种方法对 DTW 计算结果进行归一化,以便获得最优结果。

1.2 基于序列熵值加权的计算方法

通过日常生活体验很容易发现,如果两个人在晚上等休息时间经常一起出去,则其关系可能更亲密,因而他们之间的轨迹越可能相似。因此,可以使用熵值来度量用户每天活动的多样性,若某天活动越多样,则该天轨迹的相似度对总体轨迹的相似度贡献越大,进而对人们之间的关系强度贡献越大。

计算轨迹序列的熵值是为了对 DTW 计算结果进行加权,因为用户每天的轨迹序列的相似度对其总体相似度的贡献是不一样的,如果某一天

用户的轨迹序列的熵值越大,则这一天对总的相似度贡献越大。因此,使用用户每天轨迹序列熵值对用户与朋友之间每天的轨迹相似度进行加权,能够更真实地反应用户与朋友之间的关系强度(计算过程见第 2.2 节)。

1.3 基于主题模型 LDA 的计算方法

在日常生活当中,人们之间尤其是好友之间的行为模式之间具有一定的相似性,如经常在某些时间段(晚上)去一些地方(餐馆)等。基于位置的用户行为模式一方面能够反映用户在物理层次的相遇,另一方面能够在一定程度上体现用户的相似性,前文已经从社会心理学的角度阐述了相遇次数与用户关系强度的关系。文献[14]认为人们倾向于喜欢在态度、兴趣、价值观、背景和人格上与其相似的人,因此,在日常生活当中行为相似的人更可能成为朋友,而根据社会心理学的研究成果,用户的相似性对用户的关系强度也有一定的影响。为此,在通过基于用户轨迹度量用户之间关系强度的基础上,可进一步通过基于位置的用户日常行为来度量用户之间的关系强度。

LDA^[15]是一个针对离散数据集的产生式概率模型。文献[16]最先使用 LDA 主题模型发现用户的行为模式。在使用 LDA 模型发现用户基于位置的行为模式基础上,本研究进一步使用 LDA 主题模型来度量用户之间的关系强度,其核心思想如下:将每个用户每天去过的位置(语义位置或语义标签)序列视为一个句子,每个用户所有天的位置序列视为一篇文档,对所有用户所有天的位置序列使用 LDA 主题模型训练得到若干个主题。在计算两个用户之间的关系强度时,将这两个用户同一天的数据按固定长度的时间片划分;对于每个时间片内用户去过的位置,用训练好的 LDA 主题模型推断这些位置对应的主题分布;以同一时间片内,两个用户分别去过的位置对应的主题分布的余弦相似度作为这两个用户之间的关系强度(计算过程见第 2.2 节)。

2 关系强度度量模型框架

要真实全面地反映人们之间的关系强度,需从不同角度和不同层次对人们之间的关系强度进行度量,为此,提出了一个层次化的、对用户与朋友之间的关系强度进行度量并对度量结果进行投票的模型 FRSHV,其框架结构如图 1 所示。FRSHV 模型是一个三层的、能够对通过 GPS 和基站的位置数据进行处理的度量模型,其从轨迹、语

义位置以及语义标签三个层次对用户与朋友之间的关系强度进行度量,并使用集成学习的思想对三个层次度量结果进行投票,最终以投票结果作为用户与朋友之间的关系强度。

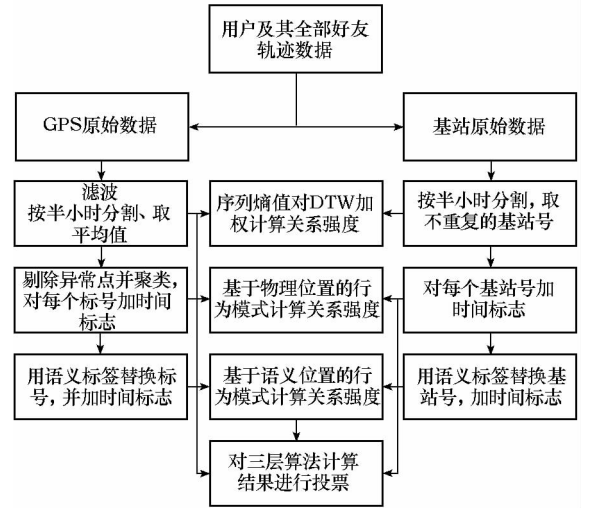


图 1 FRSHV 模型框架
Fig. 1 FRSHV model framework

在 FRSHV 模型当中,第一层度量主要针对用户的轨迹序列数据,根据不同用户轨迹序列的相似度来度量用户与朋友之间的关系强度;第二层度量主要针对用户的语义位置序列数据,考虑用户个人的基于位置的行为模式,如经常在什么时间出现在哪些位置等,根据不同用户行为模式的相似度来度量用户与朋友之间的关系强度;第三层度量主要针对用户的语义标签序列数据,物理上不同的位置可能拥有相同的语义标签,“办公室”“家”等语义概念在每个用户轨迹中都可能出现,而这些语义概念在原始数据中会表现为不同的基站号和区域号或不同的 GPS 经纬度,因此用户的语义标签数据更能体现用户群体的日常习惯,因此本层考虑的行为模式更倾向于群体的行为模式,从而根据不同用户在群体中表现出的行为模式来度量用户与朋友之间的关系强度。

2.1 GPS 及基站位置数据处理

在日常生活中,用户的位置既可以通过智能手机内嵌的 GPS 传感器获取,又可以通过对用户所处区域内的通信基站进行定位获取,基站定位更有利于用户隐私的保护。为了满足不同用户的不同需求,FRSHV 模型能够同时对 GPS 位置数据和基站位置数据进行处理。设用户集合为 U ,其中 n 表示用户个数, D_i 表示用户 u_i 采集数据的日期的集合, m_i 表示用户 u_i 采集数据的总天数。 F_i 表示用户 u_i 的全部朋友组成的集合,其中 f_i 表示

用户 u_i 的好友的个数。所有用户所有天的轨迹数据的集合为 $Trace$, 其中 $Trace_i$ 表示用户 u_i 所有天采集的轨迹序列的集合, $Trace_{i,k}$ 表示用户 u_i 在 k 这一天的轨迹序列, $n_{i,k}$ 表示用户 u_i 在 k 这一天采集的轨迹数据的条数。对 GPS 和基站表示的用户轨迹序列进行预处理时, 使用以下三种做法分别构造三层算法的输入。

2.1.1 轨迹数据处理

1) GPS 位置数据处理。首先, 对每个用户每天的数据 $Trace_{i,k}$ 进行滤波, 目的是减少数据噪声; 而后对滤波后的数据按半小时进行划分, 将用户 u_i 的每天数据 $Trace_{i,k}$ 按时间均分为 48 份, $Sep_trace_{i,k,s}$ 表示第 i 个用户第 k 天第 s 份数据; 对 $Sep_trace_{i,k,s}$ 按经纬度计算平均值, 并将用户 i 在第 k 天新的轨迹序列表示为 $Ntrace_{i,k}$, 将用户 i 所有天采集的数据 $Ntrace_i$ 作为用户 u_i 使用第一层算法计算其与全部好友关系强度的输入。

2) 基站位置数据处理。对每个用户每天的数据按半小时进行划分, 即将用户 u_i 第 k 天的数据 $Trace_{i,k}$ 按时间均分为 48 份, $Sep_trace_{i,k,s}$ 表示第 i 个用户第 k 天第 s 份数据; 对每半个小时内数据计算依次不重复的基站号序列; 再将每天 48 份数据重新拼成一个序列 $Ntrace_{i,k}$ (表示用户 i 在 k 这一天采集的全部的数据), 目的是对每天轨迹序列降维, 以降低计算的复杂度, 将用户 i 所有天的数据 $Ntrace_i$ 作为用户 u_i 使用第一层算法的输入。

2.1.2 语义位置数据处理

1) GPS 位置数据处理。采用文献[17]中的聚类方法对所有用户的轨迹数据进行聚类, 得到全部语义位置序列 Loc 。通过聚类得到用户 u_i 在第 k 天的语义位置序列 $Ltrace_{i,k}$; 用户 u_i 的全部语义位置序列表示为 $Ltrace_i$, 所有用户的所有语义位置序列表示为 $Ltrace$, 对序列 $Ltrace$ 添加对应的时间标记后记为 $LLtrace$, 训练对应的 LDA 主题模型并记为 $LLDA(K)$, K 表示主题个数。对每个用户每天的数据按半个小时进行划分, 即将用户 u_i 的每天数据 $Ltrace_{i,k}$ 按时间均分为 48 份, $Sep_trace_{i,k,s}$ 表示第 i 个用户第 k 天第 s 份数据; 对每份数据计算不重复出现的语义位置, 并对每个位置加上时间标记。用户 u_i 在第 k 天第 s 时间段语义位置序列表示为 $Tltrace_{i,k,s}$, 将用户 i 所有天的语义位置序列 $Tltrace_i$ 作为用户 u_i 使用第二层算法计算其与全部好友关系强度的输入。

2) 基站位置数据处理。将每一个基站视为

一个语义位置, 即 $Ltrace = Trace$, 其余处理与 GPS 位置数据处理完全相同。

2.1.3 语义标签数据处理

1) GPS 位置数据处理。对前文得到的序列 Loc 中每一个语义位置采用文献[17]中的方法标记其语义标签, 标记语义标签后, 用户 u_i 第 k 天的语义标签序列表示为 $Strace_{i,k}$, 用户 u_i 的全部语义标签序列表示为 $Strace_i$, 所有用户的所有语义标签序列表示为 $Strace$, 对序列 $Strace$ 添加对应的时间标记后记为 $SStrace$, 训练对应的 LDA 主题模型并记为 $SLDA(K)$, K 表示主题个数。对每个用户每天的数据按半个小时进行划分, 即将用户 u_i 的每天数据 $Strace_{i,k}$ 按时间均分为 48 份, $Sep_trace_{i,k,s}$ 表示第 i 个用户第 k 天第 s 份数据; 对每份数据计算不重复出现的语义标签, 并对每个位置对应的语义标签加上时间标记。用户 u_i 在第 k 天第 s 时间段内的语义位置序列表示为 $Tstrace_{i,k,s}$, 将用户 i 所有天的语义标签序列 $Tstrace_i$ 作为用户 u_i 使用第三层算法计算其与全部好友关系强度的输入。

2) 基站位置数据处理。计算每一个基站对应的语义标签, 其余处理与 GPS 数据处理完全相同。

2.2 关系强度计算

计算每一个用户 u_i 与其每一个朋友 u_k ($u_k \in F_i$) 之间的关系强度, 并对 F_i 中的每一个朋友, 按照其与 u_i 的关系强度大小降序排列, 使此序列中任意两个朋友与 u_i 的关系强弱顺序尽可能与实际情况一致。

1) 基于 DTW 及序列熵值加权计算用户之间的关系强度。对用户 u_i 的每一个好友 u_k , 利用 2.1.1 节中得到的 $Ntrace_i$ 和 $Ntrace_k$ 计算其轨迹序列相似度。 $Ntrace_{i,a}$ 表示用户 u_i 在第 a 天的数据, 其中 $a \in D_i$; $Ntrace_{k,b}$ 表示用户 u_k 在第 b 天的数据, 其中 $b \in D_k$ 。 $DTW(Ntrace_{i,a}, Ntrace_{k,b})$ 表示用户 u_i 在 a 这一天的轨迹和用户 u_k 在 b 这一天的轨迹的相似度, $Entropy(Ntrace_{i,a})$ 表示用户 u_i 在 a 这一天的轨迹序列的熵值。用户 u_i 和用户 u_k 的基于轨迹序列的关系强度计算方法见式(1)。DTW 计算的是距离, 距离越小相似度越大, 即该公式值越小, 则两个用户关系强度越大。

$$Ent_Dtw(u_i, u_k) = \sum_{a \in D_i, b \in D_k} S(a, b) \frac{DTW(Ntrace_{i,a}, Ntrace_{k,b})}{Entropy(Ntrace_{i,a})} \quad (1)$$

其中, 若 $a = b$, 则 $S(a, b) = 1$; 若 $a \neq b$, 则 $S(a, b) = 0$ 。

2) 基于主题模型计算用户之间的关系强度。

$Tltrace_i$ 表示用户 u_i 根据第 2.1.2 节得到的语义位置序列, $Tltrace_k$ 表示用户 u_k 根据第 2.1.2 节得到的语义位置序列。 $LLDA(K).inf(Tltrace_{i,a,p})$ 表示对 $Tltrace_{i,a,p}$ 推断得到的主题分布, 通常表示为 K 维的向量, 其中 K 表示主题的个数。基于用户语义位置的行为模式的关系强度计算方法见式(2), 其中 cos 表示余弦相似度。

$$LocLDA(u_i, u_k) = \sum_{a \in D_i, b \in D_k} S(a, b) \sum_{p=q=1}^{48} T(a, p, b, q) \cdot \cos(LLDA(K).inf(Tltrace_{i,a,p}), LLDA(K).inf(Tltrace_{k,b,q})) \quad (2)$$

其中, 若用户 u_i 在 a 这一天第 p 个时间段和用户 u_k 在 b 这一天第 q 个时间段数据均存在, 则 $T(a, p, b, q) = 1$, 否则 $T(a, p, b, q) = 0$ 。

基于用户语义标签的行为模式的关系强度计算公式与基于语义位置的关系强度计算公式相似, 见式(3)。

$$SemLDA(u_i, u_k) = \sum_{a \in D_i, b \in D_k} S(a, b) \sum_{p=q=1}^{48} T(a, p, b, q) \cdot \cos(SLDA(K).inf(Tltrace_{i,a,p}), SLDA(K).inf(Tltrace_{k,b,q})) \quad (3)$$

本研究更关注的是用户和好友 A 的关系强度大于或小于用户与好友 B 的关系强度, 因此实际计算结果应为用户与其全部好友按关系强度降序排列得到的好友序列。对于用户 u_i , 对其全部好友 F_i 中的每一个朋友 u_k 使用 $Ent_Dtw(u_i, u_k)$ 计算用户 u_i 和用户 u_k 之间的关系强度, 对 F_i 中的每一个朋友按照计算得到的关系强度降序排列得到 $E_i = \{u_{d_1}, \dots, u_{d_{f_i}}\}$; 在此基础上, 使用 $LocLDA(u_i, u_k)$ 计算用户 u_i 和用户 u_k 之间的关系强度, 并对 F_i 中的每一个朋友按照计算得到的关系强度降序排列得到 $L_i = \{u_{l_1}, \dots, u_{l_{f_i}}\}$; 最后使用 $SemLDA(u_i, u_k)$ 计算用户 u_i 和用户 u_k 之间的关系强度, 并对 F_i 中的每一个朋友按照计算得到的关系强度降序排列得到 $S_i = \{u_{s_1}, \dots, u_{s_{f_i}}\}$ 。

2.3 结果投票

采用集成学习的思想对三个层次的计算结果 E_i, L_i, S_i 进行投票, 投票规则为: 对于与用户 u_i 关系第 k 强的好友 u_{v_k} ($k \geq 1$ 且 $k \leq f_i$), 使用三个层次对应的方法分别计算得到 u_{d_k}, u_{l_k} 和 u_{s_k} 。若这三个用户都不相同, 则认为 $u_{v_k} = u_{d_k}$, 若某个用户比如 $u_{l_k} = u_{s_k}$ 出现两次及以上, 则认为 $u_{v_k} = u_{l_k}$, 最终以 $V_i = \{u_{v_1}, \dots, u_{v_{f_i}}\}$ 作为投票结果。

3 数据集及评估方法

3.1 移动数据集

在实验验证过程中, 使用 MIT 媒体实验室采集的 The Reality Mining Data 数据集^[1]。实验中使用到的信息主要包括每个用户每天由基站号组成的轨迹序列、所有用户之间的朋友关系以及各个用户的调查问卷, 同时数据集中还提供了每个基站号和区域号对应的位置的语义标签。数据集中采集的位置信息是基站信息。虽然基站定位方式的精确度比 GPS 定位方式的低, 但其更有利于用户隐私的保护, 这也是选择此数据集进行实验的主要原因之一。

在对数据集的分析过程中发现朋友关系信息表中存在如下问题: ①部分用户自己和自己是好朋友, 另外一部分用户自己和自己不是好朋友; ②某用户和另一个用户是好朋友, 而另一个用户和该用户不是好朋友。用户之间的好友关系应该满足反自反和对称。经过这样处理后, 得到好友数大于 1 的用户共有 34 个, 剔除只有一个好友的用户。在后面的实验中, 使用这 34 个用户及其全部朋友的数据来对 FRSHV 模型进行验证。

3.2 评估方法与基准

根据前文提到的社会心理学的一些研究成果, 态度、兴趣、价值观、背景和人格等方面更相似的人关系更亲密, 尤其是对生活在一起的一个群体来说, 如果在这些方面类似并且对某些问题的看法相似, 则其关系可能就更加紧密。在现实生活当中, 通常通过问卷调查方式来获得这些方面的信息, 问卷调查结果是这些方面的一种真实体现和反映, 因此, 问卷调查结果越相似的用户关系越亲密, 为此, 本文以数据集中问卷调查回答结果的相似性作为朋友之间真实的关系强度。

经过对数据集中问卷调查的仔细分析发现问卷调查中的所有问题基本上可以分为两类: 第一类问题可以用“是”或“否”来回答, 另一类问题答案多选, 但是每个选项按顺序呈现强度增强、次数增加或者次数减少。为了计算用户与朋友之间的真实的关系强度, 针对这两类问题, 采用不同的评分方法。针对第一类问题当中的每一个问题, 如果两个朋友的答案相同, 则评分为 1, 否则评分为 0; 针对第二类问题当中的每一个问题, 如果两个朋友的答案越接近, 则评分越高, 并且将评分归一化到 0~1 之间, 使得每个问题在总的关系强度评分中占有相同的权重。在完成对所有问题评分基

基础上,对所有评分进行累加求和,以此作为两个朋友之间的关系强度。依次对每个用户及其所有朋友按上述方法计算其与每个朋友之间的关系强度,并对其所有朋友的评分按降序排列,得到一个用户与其所有朋友之间的关系强度序列,以此序列作为该用户与其朋友之间真实的关系强度。在此基础上,将使用 FRSHV 模型计算出来的用户与朋友之间的关系强度序列与真实的关系强度序列进行对比,验证 FRSHV 模型的有效性。

为了度量使用 FRSHV 模型计算出来的用户与朋友之间关系强度序列 V_i 与真实的关系强度序列 G_i 的一致性,提出一种基于逆序对数的有序序列一致性度量方法。设 A 为一个有 N 个数字的有序集($N > 1$),且所有数字均不相同,如果存在正整数 i, j ,使得 $1 \leq i < j \leq N$,而 $A[i] > A[j]$,则称 $\langle A[i], A[j] \rangle$ 为 A 的一个逆序对。 A 中全部的逆序对的个数称为逆序对数。我们把序列 G_i 作为有序集,来计算序列 V_i 的逆序对数。设该用户共有 f_i 个好友,若逆序对数为 0,说明实验结果与实际结果完全一致;若逆序对数为 $\frac{f_i(f_i - 1)}{2}$,则说明实验结果恰好是实际结果的逆序。提出的有序序列一致性度量公式见式(4),其中 f_i 为用户 u_i 的全部好友的个数, K_i 为 V_i 相对于 G_i 的逆序对数。对每个用户可计算得到一个一致性评分,在此基础上,对所有用户的一致性评分取平均值,以此作为模型 FRSHV 对朋友关系强度度量有效程度的度量,见式(5)。

$$\text{score}(u_i) = 1 - \frac{K_i}{f_i(f_i - 1)/2} \quad (4)$$

$$\text{score} = \frac{1}{n} \sum_{i=1}^n \text{score}(u_i) \quad (5)$$

4 实验验证及分析

实验环境为 windows 7 64 位,4 核,3.2 GHz 主频,8 G 内存,使用 Python 编码实现。

因为用户之间的物理距离难以直接确定,所以以基站之间的距离作为用户之间的物理距离。采取如下方法来定义基站之间的距离:将每天用户手机连接过的基站视为一条基站序列,对于基站 A 和 B ,从所有用户所有天的基站序列中找到同时出现 A 和 B 的序列,计算每个序列中 A 和 B 中间不同的基站号的个数,取最小值加 1 作为基站 A 和基站 B 之间的距离。若通过上述方法能够计算出两个基站之间的距离,则称这两个基站之间的距离存在。若 A 和 B 从未在同一个基站序列中出现过,则定义

A 和 B 之间的距离为所有两个基站之间最大距离的 K 倍(K 为一个正实数参数,在后面实验中能够看到该参数对实验结果的影响)。

4.1 基于轨迹相似性计算用户之间的关系强度

通过上文对基站距离的定义,使用 DTW 以及归一化后的 DTW 计算第一层用户之间的相似度,一致性评分可通过式(4)和式(5)计算得到,上文论述到使用参数 K 定义两个不存在距离的基站的距离,不同的参数 K 以及不同方法对结果的影响见图 2。通过观察实验结果发现,取不同的 K 值会产生不同的结果,当值取得很大时候,意味着如果这两基站之间不存在距离,在实际处理过程中则认为两个基站之间距离比较大,这样会得到更理想的结果。

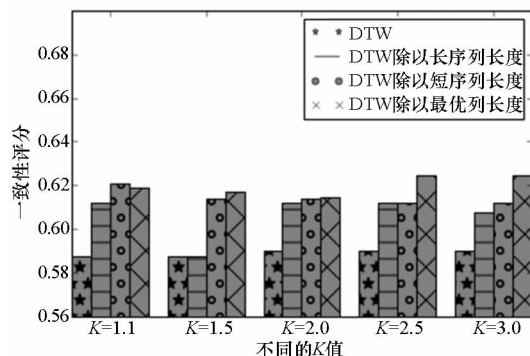


图 2 参数 K 对一致性评分结果的影响

Fig. 2 Influences of K on the consistency

在上一个实验的基础上,对 DTW 方法以及归一化的 DTW 方法使用序列熵值加权,对应 2.2 节的 E_i ,并以编辑距离^[18]计算的结果作为基准,一致性评分的实验结果见图 3。

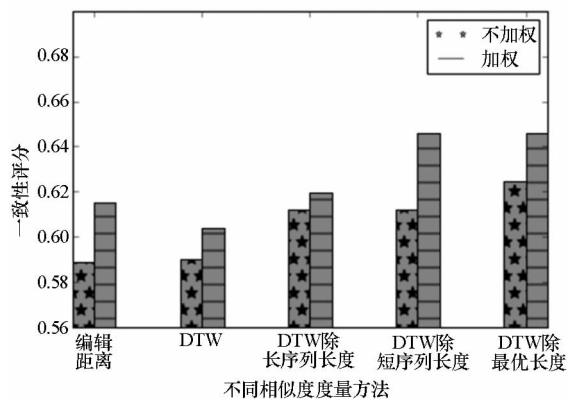


图 3 用熵值加权前后结果对比($K=2.5$)

Fig. 3 Comparison of weighted consistency and non-weighted consistency ($K=2.5$)

4.2 基于语义位置相似性计算用户之间的关系强度

在计算关系强度的过程中,使用 LDA 模型进

行推断。因为推断过程进行随机初始化,使得LDA模型的每次执行结果不一定完全相同,因此,在实验中,针对每个不同的参数值(即主题个数)执行10次,并将每次计算获得的 L_i 与 G_i 进行一致性评分。对所有用户按式(5)计算最终的一致性评分,进而取这10个一致性评分的中位数作为该参数对应的一致性评分,如图4所示。

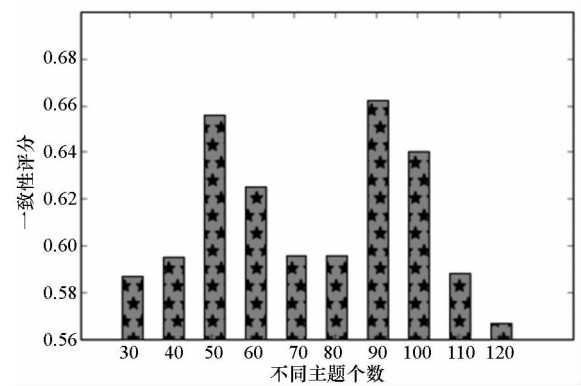


图4 使用语义位置时主题个数及对应的一致性评分实验结果

Fig.4 Influence of topic numbers to consistency using semantic location

4.3 基于语义标签相似性计算用户之间的关系强度

数据集中提供了基站号和区域号对应的位置的语义标签^[1],对所有语义标签加上时间标记,将每个带时间标记的语义标签视为单词,每天的语义标签序列视为句子,每个用户所有语义标签序列视为文档,使用所有用户的全部文档对LDA模型进行训练,其实验过程与上面的基于语义位置的实验过程一样,最后对对应的2.2节所示的 S_i 进行一致性评分。图5展示了在主题个数取不同值时所对应的一致性评分结果。

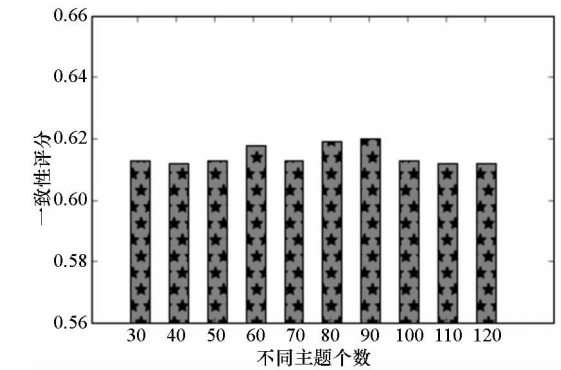


图5 使用语义标签时主题个数和对应一致性评分的实验结果

Fig.5 Influence of topic numbers to consistency using semantic label

语义标签有实际含义,以主题个数75为例,通过观察LDA模型学习到的主题,发现该模型学习得到了3个主题,如表1所示,主题1表示的是晚上在实验室或教室,主题2表示早上和晚上在家,主题3表示上午在实验室。

表1 LDA模型学习到的不同主题示例
Tab.1 Some topics of LDA learned

主题1	主题2	主题3
Tech sq_47	home_14	Media lab_17
Tech sq_46	home_15	Media lab_16
Tech sq_40	home_8	Media lab_20
Tech sq_38	home_6	Media lab_18
Tech sq_39	home_0	Media lab_19
Tech sq_42	home_44	Tech sq_17

4.4 对计算结果进行投票

上面的实验分别描述了层级模型FRSHV每一层的实验结果,在此基础上,使用前面描述的投票规则对三层中每层最好的实验结果进行投票,三层结果投票的实验结果见图6。

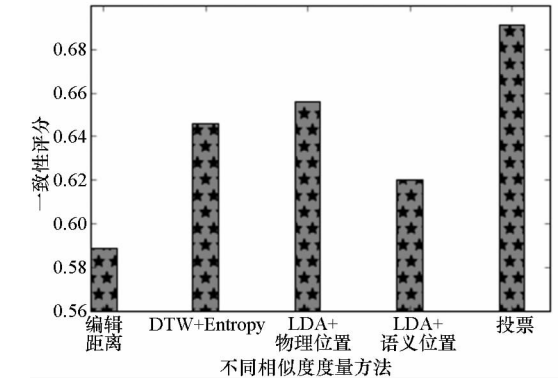


图6 投票结果及分别只使用一种方法的结果对比
Fig.6 Results comparison between vote method and simple method

通过实验结果可以发现,使用投票方法后可以更好地度量用户之间的关系强度,基于投票的方法比编辑距离一致性评分高出近10%。

5 结论

针对如何度量日常生活中人们之间的关系强度问题展开研究,提出了一个从日常轨迹、语义位置以及语义标签三个层次度量用户与朋友之间关系强度的层级模型FRSHV。采用基站数据对该模型进行了验证,观察实验结果发现基于投票的方法比编辑距离一致性评分高出近10%。下一步我们将对相关度量方法进行进一步的优化,利

用更多的消息如通话记录、短信等,进而对多种数据进行融合来度量用户之间的关系强度。

参考文献 (References)

[1] Eagle N, Pentland A. Reality mining: sensing complex social systems [J]. Personal and Ubiquitous Computing, 2006, 10(4): 255 – 268.

[2] Aharony N, Pan W, Ip C, et al. Social fMRI: investigating and shaping social mechanisms in the real world [J]. Pervasive and Mobile Computing, 2011, 7(6): 643 – 659.

[3] Wang R, Chen F L, Chen Z Y, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones [C]//Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2014: 3 – 14.

[4] Stopczynski A, Sekara V, Sapiezynski P, et al. Measuring large-scale social networks with high resolution [J]. PloS One, 2014, 9(4): e95978.

[5] Granovetter M S. The strength of weak ties [J]. American Journal of Sociology, 1973, 78(6): 1360 – 1380.

[6] Wegner D M. The illusion of conscious will [M]. Cambridge, MA, US: MIT Press, 2002.

[7] Burrows R, Nettleton S, Pleace N, et al. Virtual community care? Social policy and the emergence of computer mediated social support [J]. Information, Communication & Society, 2000, 3(1): 95 – 121.

[8] Petróczy A, Nepusz T, Bazsó F. Measuring tie-strength in virtual social networks [J]. Connections, 2007, 27(2): 39 – 52.

[9] Ma C, Cao J N, Yang L, et al. Effective social relationship

measurement based on user trajectory analysis [J]. Journal of Ambient Intelligence and Humanized Computing, 2014, 5(1): 39 – 50.

[10] Festinger L. A theory of cognitive dissonance [M]. CA, USA: Stanford University Press, 1962.

[11] Zajonc R B. Attitudinal effects of mere exposure [J]. Journal of Personality and Social Psychology, 1968, 9(2p2): 1 – 27.

[12] Itakura F, Umezaki T. Distance measure for speech recognition based on the smoothed group delay spectrum [C]//Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 1987: 1257 – 1260.

[13] Ratanamahatana C A, Keogh E. Everything you know about dynamic time warping is wrong [C]//Proceedings of Third Workshop on Mining Temporal and Sequential Data, 2004.

[14] Singelis T M. The measurement of independent and interdependent self-construals [J]. Personality and Social Psychology Bulletin, 1994, 20(5): 580 – 591.

[15] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3: 993 – 1022.

[16] Farrahi K, Gatica-Perez D. What did you do today? Discovering daily routines from large-scale mobile data [C]//Proceedings of the 16th ACM International Conference on Multimedia, 2008: 849 – 852.

[17] Yang R S, Shi D X. SASLL: a system annotating semantic label of location [C]//Proceedings of the 7th International Symposium on UbiCom Frontiers Innovative Research, Systems and Technologies, 2015.

[18] Levenshtein V I. Binary codes capable of correcting deletions, insertions, and reversals [C]//Proceedings of Soviet Physics Doklady, 1966, 10(8): 707 – 710.

多特征融合文本聚类的新闻话题发现模型*

车 蕾^{1,2}, 杨小平¹
(1. 中国人民大学 信息学院, 北京 100872; 2. 北京信息科技大学 信息管理学院, 北京 100192)

摘 要:融合新闻命名实体、新闻标题、新闻重要段落、文本语义等多特征影响,提出基于多特征融合文本聚类的新闻话题发现模型。模型根据新闻的多特征影响,提出一种多特征融合文本聚类方法。该方法针对新闻标题、新闻重要段落等特征因素构建向量空间模型及相似度算法,基于潜在狄利克雷分配模型构建主题空间模型及相似度算法,针对命名实体构建命名实体模型及相似度算法,并将三种相似度算法形成最优融合。基于多特征融合文本聚类方法,模型改进了用于新闻话题发现的 Single-Pass 算法。实验是在真实新闻数据集上开展的,实验结果表明:该模型有效地提高了新闻话题发现的准确率、召回率和综合评价指标,并具有一定的自适应能力。

关键词:新闻话题;多特征融合;潜在狄利克雷分配;向量空间模型;主题空间模型

中图分类号:TP391 文献标志码:A 文章编号:1001-2486(2017)03-085-06

News topic discovery model of multi feature fusion text clustering

CHE Lei^{1,2}, YANG Xiaoping¹
(1. School of Information, Renmin University of China, Beijing 100872, China;
2. School of Information Management, Beijing Information Science & Technology University, Beijing 100192, China)

Abstract: The news topic discovery model based on multi feature fusion text clustering was proposed fusing multi features of news, such as named entities, news headlines, important paragraphs, text semantics and so on. Based on the multi feature influence of news, a multi feature fusion text clustering method was put forward in this model. In this way, vector space model and similarity algorithm based on feature words, news headlines, important paragraphs were constructed, subject space model and similarity algorithm based on latent Dirichlet allocation were constructed, named entity model and similarity algorithm based on named entities were constructed, and those three similarity algorithms were fused optimally. Based on multi feature fusion text clustering method, the Single-Pass algorithm used in the news topic discovery was improved. Experiments were carried out on the real news data set, and the experimental results show that the model can improve the accuracy rate, recall rate and comprehensive evaluation index of the news topic discovery, and have some ability of self-adaption.

Key words: news topic; multi feature fusion; latent Dirichlet allocation; vector space model; subject space model

随着信息化的发展,互联网逐渐成为人们获取信息的一个主要途径,突发新闻事件可以在互联网上瞬间传播。如何发现新闻话题、如何追踪新闻事件的发展过程,是迫切需要解决的问题。新闻话题的内容会伴随时间发展而发生变化,新闻话题的强度也会伴随时间发展经历一个从高潮到低潮的过程。如何按时间顺序挖掘新闻集合中话题的演化过程,从而帮助用户追踪感兴趣的话题,具有实际意义。因此,话题演化研究具有现实的应用背景。话题发现是话题演化研究中的关键环节。

当前话题发现的研究主要集中在建立更好的

文本表示形式和充分利用新闻语料特征两个方面。Allan 等^[1]最先采用信息检索领域的向量空间模型(Vector Space Model, VSM)构建话题模型。Yang 等^[2]运用 Rocchio 算法对基于 VSM 的话题模型进行扩展。Nallapati^[3]提出语义语言模型。Lee 等^[4]利用增量型的方法在话题发现过程中不断提炼基于话题的特征词,给予这些特征词更大的权重,从而提高话题区分能力。王少鹏^[5]利用词频-反文档频率(Term Frequency-Inverse Document Frequency, TF-IDF)算法和潜在狄利克雷分配(Latent Dirichlet Allocation, LDA)主题模型分别计算文本的相似度,然后使用 K-means 算法

* 收稿日期:2016-02-10
基金项目:国家自然科学基金资助项目(61272513);北京市教育委员会科技计划面上资助项目(KM201511232016, SM201511232004);北京高等学校青年英才计划资助项目(YETP1503)
作者简介:车蕾(1979—),女,河南洛阳人,副教授,博士研究生, E-mail:chelei@bistu.edu.cn

进行聚类分析。马晓姝^[6]将基于 VSM 命名实体特征词的文本相似度和基于潜在主题的文本相似度进行线性结合,使用文本聚类算法进行聚类。但是这些研究缺乏新闻特征词、命名实体、新闻标题、新闻重要段落等特征对话题发现的影响。

1 LDA 模型及相关概念

话题模型本质上就是一种文本的降维表示。TF-IDF 是最早的文本降维模型,该模型的不足是无法从语义层面表示文本。随后 Deerwester 等^[7]提出了隐性语义分析 (Latent Semantic Analysis, LSA) 模型,采用矩阵的奇异值分解技术对文本进行降维,以从文本中发现隐含的语义维度,该模型的不足是没有能力处理一词多义和一义多词的问题。Hofmann^[8]在 LSA 基础上提出了概率隐性语义分析 (Probabilistic Latent Semantic Analysis, PLSA) 模型,并使用期望最大化 (Expectation Maximization, EM) 算法学习模型参数,该模型的不足是参数数量会随着文集增长而线性增长,并且产生过拟合的问题。Blei 等^[9]提出了 LDA 模型,LDA 模型既是一个概率生成模型,又是一个话题模型,该模型很好地解决了 PLSA 模型出现的问题,具有很好的泛化能力。LDA 模型在文本挖掘、机器学习等领域发挥着重要作用。

1.1 LDA 基本概念

LDA 模型是一个经典主题模型,它可以计算出每篇文档的主题概率分布。在此基础上可以利用 LDA 模型获得各个节点文本内容的主分布信息,运用协作演化的思想,将内容相似度和结构信息相融合,以提升节点相似性评定的效果^[10]。在 LDA 模型中假设文档是多个隐含主题上的混合分布,各个主题是一个固定词表上的混合分布。LDA 文档生成过程的概率模型如图 1 所示,表 1 说明了该模型中各符号的含义。

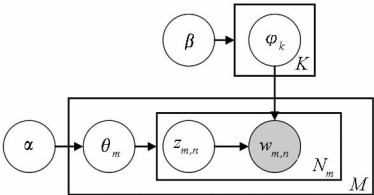


图 1 LDA 文档模型生成图

Fig. 1 LDA document model generating

LDA 模型采用 Dirichlet 分布作为概率主题模型中多项分布的先验分布。图 1 代表的概率模型如式(1)所示。

$$p(\theta,z,w|\alpha,\beta) = p(\theta|\alpha) \prod_{n=1}^N p(z_n|\theta)p(w_n|z_n,\beta)$$

(1)

表 1 LDA 模型中各符号的含义
Tab. 1 Meaning of symbols in LDA model

符号	含义
α	θ 的超参数 (先验分布)
β	φ 的超参数 (先验分布)
θ	文档 - 主题概率分布
φ	主题 - 词概率分布
M	文档数
N_m	第 m 个文档的单词数
K	主题数
$z_{m,n}$	第 m 个文档中第 n 个词的主题
$w_{m,n}$	第 m 个文档中的第 n 个词
空心圆圈	隐藏变量
实心圆圈	可观测的变量
矩形框	重复抽样过程

在 LDA 中,参数 α 和 β 是固定值,由用户事先制定。文档中的各个单词 $w_{m,n}$ 是可观测的数据。文档 - 主题概率分布以及主题 - 词概率分布,即 θ 和 φ 是隐式参数,需要通过概率推导求解。LDA 模型的参数个数只与主题数和词数有关。使用 Gibbs^[11] 采样间接估算 θ 和 φ 。

1.2 存在的问题及解决方案

LDA 主题模型可以挖掘文本内容中的潜在语义信息,但是维度较低,很难保证信息的完整性,对文本类别的区分能力不够完全。另外,对于新闻文本而言,新闻的标题和第一段,对文本类别的区分度也占有一定分量。基于 TF-IDF 的 VSM 可以发挥从词语层面对文本信息进行充分挖掘的优势。通过把 LDA 主题模型和传统的基于 TF-IDF 的 VSM 相结合,从特征词和主题两方面对文本进行聚类分析,结合两种模型的优点,弥补两种方法的缺陷。另外,也将新闻的标题、新闻重要段落、命名实体等具有明显主题特征的信息纳入模型中。该模型可以对文本信息进行充分挖掘,保证了新闻话题发现的准确度。

2 话题发现模型的构建

本模型的实施流程如图 2 所示,包括数据抓取、数据预处理、建模、计算相似度、基于改进的 Single-Pass 算法进行聚类、挖掘出相应的新闻话

题。其中,建模过程包括:VSM 特征词建模、LDA 主题建模和命名实体。计算相似度包括:计算基于特征值的文本相似度、计算基于主题的文本相似度和计算命名实体的相似度。

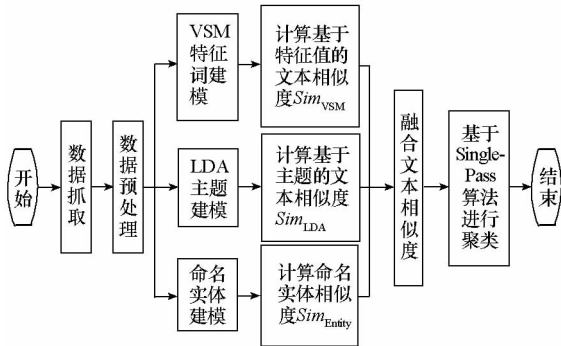


图2 多特征融合文本聚类分析流程图

Fig.2 Flow chart of multi feature fusion text clustering

2.1 多特征融合文本相似度

确定文本的相似度是进行聚类分析非常关键的一步。把基于 TF-IDF 权值策略和基于 LDA 主题模型以及基于命名实体模型的相似度进行最优线性结合,得到文本的相似度,即多特征融合文本相似度。根据计算得到的文本相似度矩阵,使用 Single-Pass 算法进行聚类分析。提取人名、地名、机构名等命名实体构建命名实体模型。

线性结合的如式(2)所示。

$$\begin{cases} sim(d_1, d_2) = \alpha sim_{VSM}(d_1, d_2) + \\ \beta sim_{LDA}(d_1, d_2) + \\ \lambda sim_{Entity}(d_1, d_2) \\ \alpha + \beta + \lambda = 1 \end{cases} \quad (2)$$

其中: α, β, λ 分别为向量空间模型的文本相似度系数、主题空间模型的文本相似度系数、命名实体模型的相似度系数。 $sim_{VSM}(d_1, d_2)$ 为两文本间向量空间模型的文本相似度; $sim_{LDA}(d_1, d_2)$ 为两文本间主题空间模型的文本相似度; $sim_{Entity}(d_1, d_2)$ 为两文本间命名实体模型的相似度。

2.1.1 向量空间模型文本相似度

不同模型对应不同的相似度计算方法,采用关键词的标准化 TF-IDF 值来衡量向量空间模型中的文本,采用余弦相似度来计算文本相似度。余弦相似度计算如式(3)所示。

$$\begin{aligned} sim_{VSM}(d_1, d_2) &= \frac{d_1 \cdot d_2}{|d_1| \times |d_2|} \\ &= \frac{\sum_{i=1}^n (a_i \times b_i)}{\sqrt{\sum_{i=1}^n a_i^2} \times \sqrt{\sum_{i=1}^n b_i^2}} \end{aligned} \quad (3)$$

其中, d_1 和 d_2 表示两个文档, a_i 表示文档 d_1 中第 i 个向量, b_i 表示文档 d_2 中第 i 个向量。

新闻模型如下: News (NewsID, NewsTitle, Content, FirstParagraph, NewsURL, Source, Date, TopicID)。新闻文本模型 New_i 中每个特征向量 w 的权重计算如式(4)所示。

$$\begin{cases} Weight(w, New_i) = \chi TFIDF(w, New_i) + \\ \lambda Count(w, NewsTitle_i) + \\ \kappa Count(w, firstParagraph_i) \\ \chi + \lambda + \kappa = 1 \end{cases} \quad (4)$$

其中: χ, γ, κ 分别为 TF-IDF 值、向量 w 在新闻标题中出现的次数、向量 w 在新闻第一段出现的次数的系数。

VSM 模型中,话题向量的每个特征向量(词语) w 的计算如式(5)所示。

$$\begin{aligned} Weight(w) &= \sum_{i=1}^{NewsCount} \left[\frac{M}{TimeDist(New_i + M)} \right] \times \\ &TFIDF(w, New_i) \times \\ &\frac{Count(w \in title)}{NewsCount} \end{aligned} \quad (5)$$

其中: $NewsCount$ 表示该话题下的新闻数目; $TimeDist(New_i + M)$ 表示第 i 个新闻的开始时间与话题的开始时间的距离, M 为调整参数; $TFIDF(w, New_i)$ 表示第 i 个新闻中出现词语 w 的权重。

考虑到文本长度对权值的影响,需要对特征权值公式做归一化处理,将各权值规范到 $[0, 1]$ 之间,如式(6)所示。

$$W_{ik} = \frac{tf_{ik} \times \ln\left(\frac{N}{n_k} + 0.01\right)}{\sqrt{\sum_{k=1}^n [(tf_{ik}) \times \ln\left(\frac{N}{n_k} + 0.01\right)]^2}} \quad (6)$$

2.1.2 主题空间模型文本相似度

采用服从 Dirichlet 分布的主题概率向量来衡量 LDA 主题模型中的文本,并用延森 - 香农 (Jensen-Shannon, JS) 距离函数来计算文本概率向量的相似度。

$p = (p_1, p_2, \dots, p_k)$ 到 $q = (q_1, q_2, \dots, q_k)$ 的距离定义如式(7)所示。

$$D_{js}(p, q) = \frac{1}{2} \left(\sum_{j=1}^k p_j \ln \frac{2p_j}{p_j + q_j} + \sum_{j=1}^k q_j \ln \frac{2q_j}{p_j + q_j} \right) \quad (7)$$

其中, p, q 为主题概率分布。

采用 LDA 模型对文本向量进行建模,并使用 Gibbs 抽样法对建模后的文本向量矩阵进行求解,得到文本 - 主题矩阵和主题 - 词矩阵,从而获

取每个文本的概率向量。主题空间模型中,每个主题的概率向量计算如式(8)所示。

$$\text{主题的概率向量} = \left(\frac{\sum_{i=1}^{\text{NewsCount}} d_{11}}{\text{NewsCount}}, \frac{\sum_{i=1}^{\text{NewsCount}} d_{12}}{\text{NewsCount}}, \dots, \frac{\sum_{i=1}^{\text{NewsCount}} d_{1t}}{\text{NewsCount}} \right) \quad (8)$$

2.1.3 命名实体相似度

命名实体部分使用的是 Jaccard 来计算其相似度,如式(9)所示。

$$\text{sim}_{\text{Entity}}(d_1, d_2) = \frac{\text{count}(\{e \in d_1\} \cap \{e \in d_2\})}{\text{count}(\{e \in d_1\} \cup \{e \in d_2\})} \quad (9)$$

其中, d_1 、 d_2 分别表示两个实体, e 表示实体里面的每个对象。

每个话题下的命名实体就是该话题下所有新闻的命名实体的并集。

2.2 多特征融合文本聚类算法

文本聚类的主要方法可分为 5 类^[7]: 划分方法、层次方法、基于密度的方法、基于网格的方法和基于模型的方法。基于划分的聚类算法需要对 K 个分组进行初始化, 并通过迭代方法将数据聚合到分组中, 使得每次得到的分组方案较前一次好; 基于层次的聚类方法通过对数据集按照某种指定的方法进行层次划分, 直到满足收敛或者满足某种条件时停止; 基于密度的聚类方法是基于密度的, 而其他的算法基于各种不同的距离计算方式, 其克服了基于距离的算法只能发现一定距离内的类簇的局限性; 基于网格的聚类方法基于网格结构划分文档集合, 然后在网格上进行聚类; 基于模型的聚类方法依据建立的数学模型, 对给定的文档数据与该数学模型进行拟合。

Single-Pass 算法是话题检测的一种著名算法, 实际是 K 最近邻 (K -Nearest Neighbor, KNN) 的一种应用, 为 KNN 算法^[12]。Single-Pass 算法常被用于 Web 新闻话题发现, 其算法的优劣与新闻的报道时间有关, 并且该算法是一种增量的聚类算法, 因此适合于 Web 话题检测。基于多特征融合相似度改进了新闻话题发现算法 Single-pass, 改进算法的步骤如图 3 所示。提出的多特征融合文本聚类算法可以应对实时新闻报道, 并对其进行动态聚类。

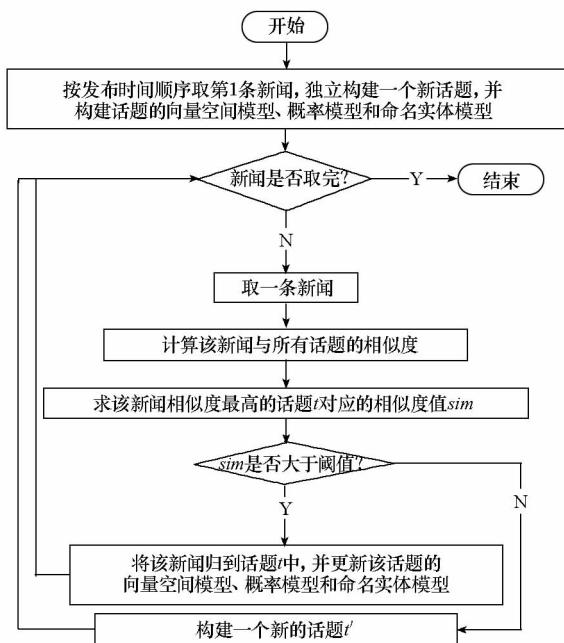


图 3 基于多特征融合文本聚类的新闻话题发现步骤

Fig. 3 Steps of new topics discovering based on multi feature fusion text clustering algorithm

3 实验设计及结果分析

3.1 实验环境及数据

实验是在 CPU 为 Intel(R) Core(TM) i5 - 3320M CPU @ 2.60 GHz、内存为 8 GB、操作系统为 Windows 8.1 专业版、处理器基于 X64 的 PC 机上运行的。算法基于 Java 语言实现。

实验数据是 2014 年 2 月至 2015 年 4 月从新闻网站的专题版块抓取的 20 个专题的相应新闻, 共 3009 条新闻。首先对这些新闻数据进行预处理, 包括分词、去停用词、识别命名实体。然后构建新闻的特征向量表, 并结合词频、新闻标题、新闻第一段等多特征计算特征向量的权重。为下一步的实验分析做好准备。

3.2 性能评价指标

为了能够对话题检测的效果进行有效的评价, 话题检测与跟踪 (Topic Detection and Tracking, TDT) 提出了评价标准, 包括: 准确率、召回率、漏报率、误报率和综合评价指标 (F1-Measure)。准确率、召回率、综合评价指标越高越好, 漏报率和误报率越低越好, 预测主题数目越接近实际主题数目越好。其中:

$$\text{准确率} = \frac{\text{识别出的新闻数目}}{\text{新闻总数目}}$$

$$\text{召回率} = \frac{\text{识别出的关于某个话题的新闻数目}}{\text{语料库中描述该话题的新闻数目}}$$

漏报率 = $\frac{\text{没有识别出的与某话题相关的新闻数目}}{\text{语料库中描述该话题的所有新闻数目}}$

误报率 = $\frac{\text{对某话题识别错误的新闻数目}}{\text{语料库中与该话题不相关的新闻数目}}$

综合评价指标 = $\frac{2 \times \text{准确率} \times \text{召回率}}{\text{准确率} + \text{召回率}}$

3.3 实验结果与分析

实验分析主要包括如下内容:通过多次实验结果的比对,确定令性能评价指标达到最优的相似度阈值;当阈值确定后,通过多次实验结果的比对,确定令性能评价指标达到最优的特征向量权重各影响因素因子的取值;通过 LDA 实验,确定 LDA 最优主题数目;最后,3.3.4 节基于确定好的各个参数以及相同实验数据,分别开展仅采用 VSM 方法、VSM-多特征 (Multi Feature, MF) 方法、LDA 方法、VSM-LDA 和多特征融合 (Multi Feature Fusion, MFF) 方法的实验,通过比对各实验结果,证明基于多特征融合方法的文本聚类模型的有效性和优势。所有最优参数值或阈值都是由系统自动根据具体数据源计算获得。

3.3.1 相似度阈值的选择

将实验结果数据首先按 |预测主题数目 - 实际主题数目| (即差的绝对值) 的升序排序,再按 F1-Measure 的降序排序,排在第一位的数据就是综合评价最优的。图 4 显示了不同相似度阈值下的实验对比结果,从中可以分析出,阈值取 0.4 时综合评价最优的。

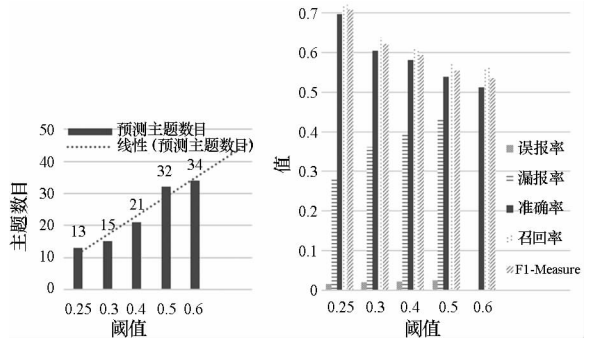


图 4 不同阈值下的主题数目估计情况、误报率、漏报率、召回率、准确率和 F1-Measure

Fig. 4 Subject number estimation, false positive rate, false negative rate, recall rate, accuracy rate and F1-Measure under different thresholds

3.3.2 特征向量权重各影响因素权重的选择

影响因子的特征向量权重如式 (4) 所示,为找到最优特征向量权重各影响因素的权重,实验遍历 3 个影响因素的所有组合的可能。图 5 显示了特征向量权重各影响因素不同权重组合的性能

评价指标,该图显示的仅是综合评价较好 (即预测主题数目接近实际主题数目, F1-Measure 较高) 的一些影响因素权重组合。综合 F1-Measure 和主题数目考虑,当 $\chi = 0.6, \gamma = 0.3, \kappa = 0.1$ 时性能评价指标最好。从该实验可以看出,考虑标题和第一段对主题的影响相比仅考虑传统 TF-IDF 对主题的影响的评价效果要好。

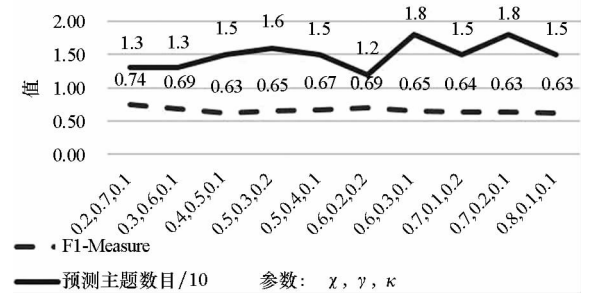


图 5 特征向量权重各影响因素不同权重组合的性能

Fig. 5 Performance of different weight combinations of each effected factors of feature vector weight

3.3.3 LDA 最优主题数 T 的选择

为确定使 LDA 算法达到最优性能评价指标所对应的相似度阈值和主题数,实验仅用 LDA 算法计算相似度,遍历了 0.001 至 0.5 范围的相似度阈值和 10 至 150 之间的 LDA 主题数目,将实验结果数据首先按 |预测主题数目 - 实际主题数目| (即差的绝对值) 的升序排序,再按 F1-Measure 的降序排序,筛选出前 n 条实验结果,筛选结果如图 6 所示。综合 F1-Measure 和主题数目, LDA 主题数目为 50 时,预测主题数目为 20, F1-Measure 为 0.807 180,性能综合评价最好。

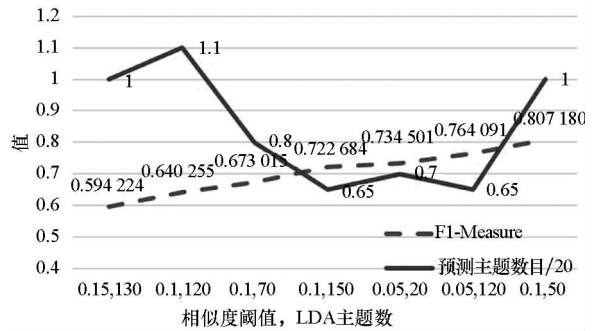


图 6 不同相似度阈值不同 LDA 主题数的性能

Fig. 6 Performance of different similarity thresholds and different LDA subject

3.3.4 多特征融合模型

为确定使多特征融合模型达到最优性能评价指标,实验遍历了式 (2) 中的 α, β 和 λ 所有可能组合,从实验结果中筛选出预测主题数在 18 至 30 之间的记录,筛选结果如图 7 所示。综合 F1-

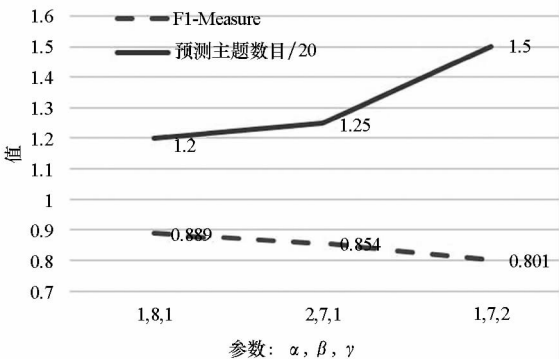


图 7 多特征融合模型中各因素影响因子的选择

Fig.7 Choice of effected factors in multi feature fusion model

Measure 和主题数目考虑, α 、 β 和 λ 分别取 1,8,1 的时模型性能评价指标最好。表 2 显示了多特征融合模型 (MFF) 与其他模型 (VSM、VSM-MF、LDA 和 VSM-LDA) 的评价指标比对情况。其中,VSM、LDA、VSM-LDA 模型都有同行做过相关研究。从表 2 中可以看出,结合了多特征的 VSM-MF 比单纯的 VSM 的性能要优;虽然 LDA 的性能要优于 VSM 和 VSM-MF,VSM-LDA 的性能优于前 3 个,但是提出的将 VSM、LDA 和多特征因素结合起来的 多特征融合模型 MFF 的性能相比之下是最优的。

表 2 多特征融合模型与其他模型的评价指标比对

Tab.2 Comparison of evaluating index of multi feature fusion model and other models

模型	准确率	召回率	漏报率	误报率	F1-Measure
VSM	0.581	0.608	0.392	0.022	0.594
VSM-MF	0.643	0.664	0.336	0.019	0.654
LDA	0.830	0.785	0.215	0.009	0.807
VSM-LDA	0.816	0.828	0.172	0.010	0.822
MFF	0.901	0.878	0.122	0.005	0.889

4 结论

本文基于 LDA 模型提出了一种改进的多特征融合文本聚类的新闻话题发现模型,该综合考虑新闻的各组成部分、新闻的词频、新闻的语义等多特征,提高了 Web 新闻话题发现的准确性,并能更好地发现新主题。实验表明,改进模型具有更优的主题检测效果。由于实验中的所有最优参数值或阈值都是由系统自动根据具体数据源计算获得,所以模型具有很好的自适应能力。

在 Web 新闻话题发现的研究方面还存在很多问题需要进一步探讨。例如:目前多特征是按

照线性方式进行结合的,非线性的结合模型是否会更好地提高话题发现的准确度有待研究。

参考文献 (References)

[1] Allan J, Papka R, Lavrenko V. On-line new event detection and tracking [C]//Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1998: 37 – 75.

[2] Yang Y, Ault T, Pierce T, et al. Improving text categorization method for event tracking[C]//Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2000: 65 – 72.

[3] Nallapati R. Semantic language models for topic detection and tracking[C]//Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, 2003: 71 – 81.

[4] Lee C H, Wu C H, Chien T F. BursT: a dynamic term weighting scheme for mining microblogging messages [C]// Proceedings of the 8th International Symposium on Neural Network, 2011: 548 – 557.

[5] 王少鹏. 基于 LDA 的文本聚类在网络舆情分析中的应用研究 [J]. 山东大学学报: 理学版, 2014, 49 (9): 129 – 134.

WANG Shaopeng. Research of the text clustering based on LDA using in network public opinion analysis[J]. Journal of Shandong University: Natural Science, 2014, 49(9): 129 – 134. (in Chinese)

[6] 马晓姝. 基于 LDA 模型的新闻话题发现研究 [D]. 长春: 东北师范大学, 2014.

MA Xiaoshu. News topic discovery research based on the LDA model[D]. Changchun: Northeast Normal University, 2014. (in Chinese)

[7] Deerwester S, Dumais S, Furnas G W, et al. Indexing by latent semantic analysis[J]. Journal of the American Society of Information Science, 1990, 41(6): 391 – 407.

[8] Hofmann T. Probabilistic latent semantic indexing [C]// Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1990: 50 – 57.

[9] Blei D M, Ng A Y, Jordan M I, et al. Latent dirichlet allocation[J]. The Journal of Machine Learning Research, 2003, 3: 993 – 1022.

[10] 胡艳丽, 白亮, 张维明. 网络舆情中一种基于 OLDA 的在线话题演化方法 [J]. 国防科技大学学报, 2012, 34(1): 150 – 154.

HU Yanli, BAI Liang, ZHANG Weiming. OLDA-based method for online topic evolution in network public opinion analysis[J]. Journal of National University of Defense Technology, 2012, 34(1): 150 – 154. (in Chinese)

[11] Griffiths T L, Steyvers M. Finding scientific topics [J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(S1): 5228 – 5235.

[12] 吴少凯. 基于桶的二次聚类新闻热点话题挖掘及应用 [D]. 广州: 华南理工大学, 2013.

WU Shaokai. Mining and application of hot news topics by bucket-based quadratic clustering [D]. Guangzhou: South China University of Technology, 2013. (in Chinese)

用卷积神经网络分类最大稳定极值区域实现汉字区域定位*

张鹏伟, 张伟伟

(信息工程大学 密码工程学院, 河南 郑州 450001)

摘要:获取对应笔画级连通区的最大稳定极值区域,实施形态学闭操作融合相距较近的最大稳定极值区域,融合后最大稳定极值区域对应的单个汉字区域;利用灰度共生矩阵描述最大稳定极值矩形区域的纹理信息,将其作为卷积神经网络的输入,卷积神经网络对最大稳定极值区域进行分类,过滤非汉字部分;利用最大稳定极值区域颜色直方图的 Bhattacharyya 距离等特征对最大稳定极值区域进行聚类,同一类最大稳定极值区域组合得到汉字文本候选区域;再次利用卷积神经网络对候选文本区域进行分类,过滤非文本部分,剩余的就是定位到的汉字文本区域。实验结果表明,该算法对于汉字区域定位具有良好的效果。

关键词:汉字区域定位;最大稳定极值区域;卷积神经网络;深度学习;灰度共生矩阵

中图分类号:TP391.4 **文献标志码:**A **文章编号:**1001-2486(2017)03-091-06

Scene Chinese text localization by convolutional neural network classifying maximum stable extremal regions

ZHANG Pengwei, ZHANG Weiwei

(School of Cryptography Engineering, Information Engineering University, Zhengzhou 450001, China)

Abstract: Firstly, the MSERs (maximum stable extremal regions) which corresponded to Chinese strokes was extracted. The morphological close operation was used to connect the nearby MSERs. The fused MSER corresponded to Chinese characters. Gray level co-occurrence matrix was used to describe the textural characteristics of the fused MSER rectangle. They were the input of CNN (convolutional neural network). The MSER rectangles were classified by CNN in order to filter non-Chinese character rectangle. Then, Chinese text candidates were constructed by clustering MSER rectangles based on the features such as the color histogram Bhattacharyya distance of MSER rectangles. CNN was reused to classify Chinese text candidates to filter non-Chinese text clusters. Finally, the rectangle of the remaining clusters was the Chinese text regions of natural scene image. Experiment shows that the proposed algorithm is desirable in localizing the Chinese text in natural scene images.

Key words: Chinese text localization; maximum stable extremal region; convolutional neural network; deep learning; gray level co-occurrence matrix

自然场景图像除包含丰富的色彩、形状、图案等物体视觉信息外,还可能包含大量的文本信息,比如书籍封面标题、单位名称、商店名称、道路路牌、交通指示牌、街道名称、建筑物门牌号、广告牌上的文字等。这些文本信息对于基于内容的图像检索、场景理解和智能交通等应用具有重要价值,从自然场景图像中自动提取文本信息已成为研究的热点^[1]。

自然场景中文本的自动提取面临着许多困难:文本存在于自然场景图像的任意位置,且与背景往往混为一体;拍摄角度多种多样,字体纷繁复杂,透视形变严重。其中,确定文本在自然场景中的位置,即文本区域定位,是场景文本自动提取的前提和基础。文本区域定位的方法主要分为两类:一是基于滑动窗口的方法^[2-4],采用文字区域分类器扫描整个场景图像,时间复杂度较高;二是基于连通区域的方法^[5-9],认为文本是具有相近的颜色和亮度的连通区域,连通区域被作为文本的候选区域。此类方法中,最大极值区域(Maximally Stable Extremal Regions, MSER)被广泛采用^[9]。

基于连通区域的文本定位方法一般对英文文本比较有效,这是因为除*i j*外的大部分英文字符直接对应一个连通区域;对于中文文本,汉字区域通常由多个连通区域构成,其定位问题更加复杂。为此,刘晓佩、潘娜等采用小波变换捕捉汉字特性^[10-11],孙巧榆采用视觉关注模型获取显著图掩膜确定中文区域^[12],徐琼等提取候选区域的方向梯度直方图金字塔(Gabor Pyramid of Histogram of Orientation Gradients, PHOG-Gabor)特征并采用

* 收稿日期:2016-01-07

基金项目:国家 863 计划资助项目(20157011012)

作者简介:张鹏伟(1978—),男,山西偏关人,工程师,博士研究生,E-mail:zhang_pw@126.com

提升树算法确定文本区域^[13]。

课题组之前基于汉字特点设计了初步的汉字文本区域定位算法^[14],本文在其基础上,以提取 MSER 区域为基础,利用卷积神经网络过滤非文本 MSER 区域为核心,给出了一种改进的汉字区域定位算法。

1 汉字特点分析与定位思路

假设自然场景中的汉字文本采用印刷体汉字。相对于手写体,印刷体汉字的形体结构清晰,如图 1 所示。汉字包括了笔画、部首等两个层次。笔画居于汉字结构最低层次,是指汉字书写时不间断地一次写成的一个线条,笔画区域是连通区;部首处于汉字结构的中间层次,部首包括若干笔画,笔画要么连通,即使不连通间距也较近;汉字是中文文本基本单元,除独体字外,每个汉字都由若干个不连通的部首构成,但与汉字间的间距相比,不连通的部首间的间距也相对较小。

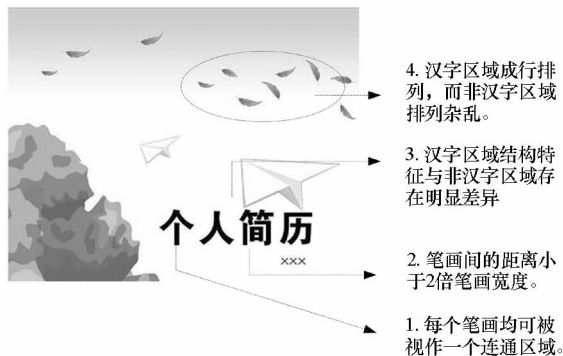


图 1 场景图像中汉字的特点

Fig. 1 Chinese characters in scene images

基于场景图像汉字中连通区域的距离关系,本文定位算法框架如图 2 所示。首先,提取 MSER 区域,它对应笔画级的连通区,通过形态学

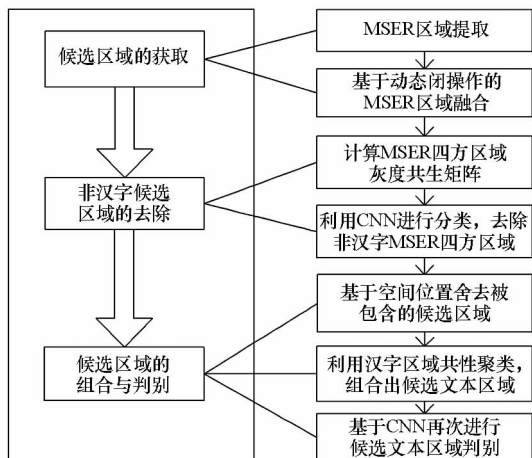


图 2 定位算法框架

Fig. 2 Chinese text localization algorithm flowchart

闭操作拓展连通区范围,使融合后的 MSER 区域尽可能对应单个汉字区域,将其作为汉字的候选区域;然后,用纹理信息作为 CNN 输入,CNN 分类融合后 MSER 区域,排除候选区域中的非汉字部分;最后,基于同一区域汉字具有共性的特点进行聚类,组合出汉字所在区域,利用 CNN 再次进行甄别,最终保留的区域就是最终定位到的自然场景汉字文本区域。

2 候选区域的获取

2.1 提取 MSER 区域

极值区域是内部像素点值要比外部像素点值低(或者高)的区域^[15]。假设对灰度图像进行二值化,二值图像中黑色区域对应的像元集合就是极值区域。当二值化的灰度阈值从 0 依次变大到 255 时,极值区域的面积将逐渐扩大,类似于水面上升,旧的极值区域被包含到新的极值区域里面,当阈值为 255 时,整个图像成为极值区域。

MSER 区域是指在某个灰度阈值 i 的时候,区域像元数量变化最小的极值区域^[15]。设 $Q_1, \dots, Q_{i-1}, Q_i, \dots$ 为一系列由于灰度阈值升高而产生的相互包含极值区域,即 $Q_i \subset Q_{i+1}$ 。 Δ 表示微小的灰度变化,当且仅当区域变化率 $Q(i) = |Q_{i+\Delta} - Q_{i-\Delta}| / |Q_i|$ 在 i 处取得局部极小值时,极值区域 Q_i 成为 MSER 区域。

MSER 区域是用不同灰度阈值对图像进行二值化时得到的最稳定的区域,区域内和区域外反差较大,因而一般具有明显的轮廓,此性质与自然图像中的汉字或汉字笔画区域比较吻合,因此 MSER 区域适合作为汉字初步的候选区域。

2.2 提取融合的 MSER 区域

大部分英文字符(除 i, j 外)都是“一笔画”的结构样式,对应一个完整的 MSER 区域,如 a, b, c, d 等。而汉字的组成单位是笔画,多个笔画先构成部首,部首再组合成汉字。部首中的笔画有相互连通的,如“扌”“女”“亻”“勹”等,也有互不连通的,如“彳”“彳”“彳”等,此外,偏旁部首间一般互不连通。一个汉字往往不能对应单个 MSER 区域。

汉字一般具有四方性,这使得单个汉字中不连通的笔画间距离都很近,而汉字间的距离往往显著大于汉字内非连通的笔画间距。形态学中的闭操作能够消除狭窄的间断和细长的鸿沟,消除小的孔洞,并填补轮廓线中的断裂。因此,选择闭操作将同一汉字内的多个笔画 MSER

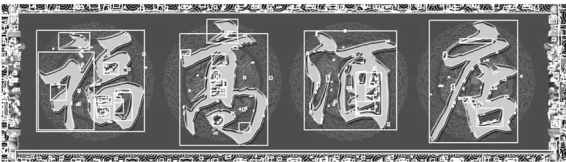
区域融合在一起。结构元及其参数影响融合的程度,结构元参数过小,导致融合不足,单个笔画 MSER 区域依然存在;参数过大,可能导致相邻汉字的笔画区域相互融合。基于汉字一般都是方块字的特点,采取方形作为结构元。对于图片汉字来说,笔画间的距离近似于笔画宽度,且小于相邻汉字的距离。用笔画宽度作为结构元参数,一般能够较好地进行同一汉字多个 MSER 区域的融合。

综上,为了进行汉字区域定位,首先提取图像中所有的 MSER 区域^[15],然后进行闭操作进行 MSER 区域融合。融合后的 MSER 区域将对应单个汉字,但轮廓上具有不规则性,为了便于描述和分析汉字的四方特性,最后沿水平方向和垂直方向提取包络 MSER 融合区域的四方区域(下面简称为 MSER 四方区域,下面如不单独声明,MSER 四方区域均指融合后的 MSER 四方区域),用于下一步的汉字区域定位处理。图3给出了提取融合后 MSER 四方区域的一个例子,图3(a)是含有汉字的场景图像,图3(b)中各个矩形框是提取出融合前的 MSER 四方区域,图3(c)中各个矩形框是提取出融合后的 MSER 四方区域,从“高”字可见,通过闭操作将“高”字头上的点融合进来了。



(a) 含有汉字的场景图像

(a) Example for scene images with Chinese character



(b) 闭操作融合前 MSER 四方区域(各个矩形框)

(b) Rectangle region of MSER before close operation



(c) 闭操作融合后 MSER 四方区域(各个矩形框)

(c) Rectangle region of MSER after close operation

图3 融合后的 MSER 四方区域示例

Fig.3 Examples for fused MSER rectangle

3 候选区域中非汉字部分的去除

3.1 计算 MSER 四方区域的灰度共生矩阵

上面提取出的 MSER 四方区域大多正好包含一个汉字,也可能包含多个汉字;但仍可能仅包含汉字的一个部分(如图3(c)中“高”字中的小矩形框),或根本不对应汉字区域(如图3(c)中的上下边框部分)。包含单个汉字和多个汉字的 MSER 四方区域在纹理上具有一定相似性,可将其视为一类,记为 H_0 类。该类与对应非文本区域和汉字局部的 MSER 四方区域(这两类 MSER 四方区域记为 H_1 类)在纹理上应有较大差异,可利用纹理特征对 H_0 类和 H_1 类 MSER 四方区域进行分类,目的是滤掉 H_1 类 MSER 四方区域。

灰度共生矩阵反映不同像素相对位置的空间信息,在一定程度上反映了纹理图像中各灰度级在空间上的分布特性,是纹理分析中最经常采用的特征之一。设 I 为一幅灰度图像,其大小为 $M \times N$,灰度共生矩阵 P 定义为:

$$P(i,j) = \# \{ [(x,y),(x+a,y+b)] | I(x,y) = i, I(x+a,y+b) = j \} / (M \cdot N) \quad (1)$$

其中, $\#$ 表示取集合元素数量; $i, j \in \{0, 255\}$; a, b 指示了像素位置的差异。灰度共生矩阵统计了两个像素点位置的联合概率分布,是图像灰度变化的二阶统计度量,也是描述纹理结构性质的基本函数。

对提取出的每个 MSER 四方区域同样可以计算灰度共生矩阵 P ,并且无论 MSER 四方区域是长方形的(区域中包含多个汉字)还是类正方形(区域中对应单个汉字),均可以得到相同维数(256×256 维)的灰度共生矩阵 P ,便于分类器的处理。因此,下面将 P 作为分类器的输入,利用分类器的输出判别 MSER 四方区域属于 H_0 类还是 H_1 类。

3.2 用卷积神经网络分类 MSER 四方区域

当获得了每个待分类的 MSER 四方区域灰度共生矩阵 P 后, $i, j \in \{0, 255\}$, P 的维数为 256×256 ,对一般的分类器而言,往往需要进一步提取灰度共生矩阵的高阶统计量以减少维数,但这往往导致了信息的丢失。最近,CNN 在图像分类上取得了巨大成功^[16],它可以直接处理高维的图像矩阵数据,而不需要先进性特征提取。因此,构建输入为 256×256 维数据的 CNN,用于对 MSER 四方区域的灰度共生矩阵进行是 H_0 类和 H_1 类的判别。

参考 AlexNet^[16] 设计 CNN,但为了提高训练

速度,仅采用了 6 层结构,如图 4 所示。第一层为卷积层,卷积核尺寸为 25×25 ,卷积步长为 1,即用 256×256 灰度共生矩阵卷积 25×25 的卷积核,取卷积结果的 valid 部分并进行非线性运算 ReLU,得到 $(256 - 25 + 1) \times (256 - 25 + 1)$ 维的特征 map,共 100 种卷积核,将得到 100 个 232×232 维特征 map,其中的每一个值对应一个神经元;第二层为 Max 池化层,其主要作用是在保证纹理细节被保留的前提下进行降维,为了降低运算量,取池化窗口大小为 8×8 ,步长为 8,池化处理后得到 100 个 29×29 维的特征 map;第三层为卷积层,卷积核尺寸为 $100 \times 6 \times 6$,共 48 种卷积核,卷积后得到 48 个 24×24 维的特征 map;第四层为 Max 池化层,池化窗口大小为 2×2 ,步长为 2,池化后将降维为 48 个 12×12 的特征 map;第五层为全连接层,共有 1000 个神经元,第六层为输出层,共有两种输出,分别对应 H_0 类和 H_1 类。

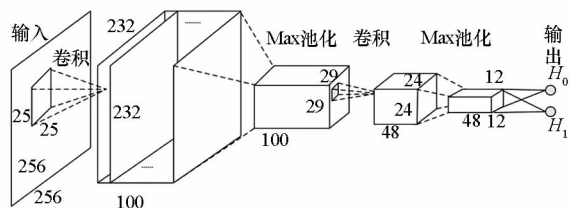


图 4 CNN 结构

Fig. 4 CNN architecture

CNN 神经元参数的有效训练是取得良好分类性能的关键,训练一方面在于训练算法,另一个方面要有大量有效的两类样本数据。训练算法方面,采用随机梯度下降(stochastic gradient descent)迭代算法^[16]更新神经元参数,损失函数采用交叉熵代价函数;训练样本数据方面,提取大量 H_0 类和 H_1 类的 MSER 区域,分别计算其灰度共生矩阵作为两类样本。

4 候选区域的组合与文本区的确定

4.1 基于空间位置对候选区域取舍

经过 CNN 分类后,得到许多候选区域(即判定为 H_0 类的 MSER 区域),这其中的一些区域仍然可能存在包含关系,如图 3(c)中“福”字对应的四方候选区域就包含了偏旁对应的小候选区域,需要舍去小区域,保留大区域,大区域更可能对应整个汉字。删减步骤如下:

步骤 1:按候选区域左上角的坐标位置,从左到右、从上到下的顺序对所有区域排序,形成候选区域队列。

步骤 2:取出队头两个候选区域,若二者存在完全包含关系,即队头区域的右下角坐标位置在第二个区域右下角的右下方,则保留队头区域,第二个区域从候选区域队列中去除(即合并到大队头区域中),队列长度减 1,然后重复步骤 2。

步骤 3:若二者不存在完全包含关系,保持队列长度不变,仅将队头指向原队头下面的第二个区域,然后判断从新队头开始的队列中是否还有两个区域(含队头区域),若是(即仍可以比较)则转到步骤 2,否则退出,说明已经取舍完毕。

4.2 对候选区域进行聚类

取舍完成后的候选区域将不会存在包含关系,但仍有可能存在交叠关系,或距离很近。场景图像中的文本通常有多个汉字(字符),这些距离很近的候选区域往往正对应了同一文本区的多个汉字或汉字组成部分,需要将这些区域进行组合,进而得到完整汉字文本区域。

场景图像中属于同一文本区的汉字一般色彩纹理统一,即具有相近属性且距离较近,因此可依据色彩相近程度以及距离关系对候选区域进行聚类,聚类后属于同一类的候选区域将进行组合。

聚类算法采用 Yin 的 single-link 聚类算法^[9]。在 single-link 算法中,需要计算两两候选区域的特征距离或特征差,聚类性能很大程度上依赖于选择的特征上,也就是对候选区域的描述方法上。本文沿用四种特征距离:空间距离、宽高差、顶部和底部对齐程度、笔画差^[9]。在此基础上采用前续工作^[14]中的颜色直方图特征矢量,增加 Bhattacharyya 距离^[17]作为区域色彩相近程度的度量而不是简单的欧式距离度量。颜色直方图估算了各种色彩出现的概率分布,Bhattacharyya 距离是一种衡量两个概率分布的相似程度的距离度量,两者结合起来能更为准确地刻画候选区域汉字颜色是否相近这一特点,因而有助于提高聚类性能。

颜色直方图计算方法如下:首先把候选区域的红绿蓝(Red Green Blue, RGB)色彩空间数据转换到色度/饱和度/亮度(Hue Saturation Value, HSV)空间,然后按照人的视觉分辨能力,把色调 H 空间分成 7 份,饱和度 S 和亮度 V 空间分成 2 份,得到:

$$H_i = \begin{cases} 0 & h \in (330, 360] \cup [0, 22], i = 1 \\ 1 & h \in (22, 45], i = 2 \\ 2 & h \in (45, 70], i = 3 \\ 3 & h \in (70, 155], i = 4 \\ 4 & h \in (155, 186], i = 5 \\ 5 & h \in (186, 278], i = 6 \\ 6 & h \in (278, 330], i = 7 \end{cases} \quad (2)$$

$$S_j = \begin{cases} 0 & s \in [0, 0.65], j = 1 \\ 1 & s \in (0.65, 1], j = 2 \end{cases} \quad (3)$$

$$V_k = \begin{cases} 0 & v \in [0, 0.7], k = 1 \\ 1 & v \in (0.7, 1], k = 2 \end{cases} \quad (4)$$

统计 (H_i, S_j, I_k) 中各个值在候选区域 $area_q$ 出现的相对次数,即:

$$P_q(i, j, k) = \# \{ \delta_{m,n} | h(m, n) = i, s(m, n) = j, s(m, n) = k, \delta_{m,n} \in area_q \} / \# area_q \quad (5)$$

其中, $\forall \delta_{m,n} \in area_q, area_q$ 为第 q 个候选区域; $\delta_{m,n}$ 为候选区域中坐标位置为 (m, n) 的像素点; $\# area_q$ 表示取候选区域中像素的总个数; $(i, j, k) \in (H_i, S_j, I_k)$ 。 $P_q(i, j, k)$ 就是对候选区域 $area_q$ 提取的颜色直方图。

对两个候选区域 $area_p$ 和 $area_q$ 得到的颜色直方图 $P_p(i, j, k)$ 、 $P_q(i, j, k)$, 其 Bhattacharyya 距离的计算公式为:

$$J_B(P_p, P_q) = - \ln \sum_{i,j,k} [P_p(i, j, k) \cdot P_q(i, j, k)]^{1/2} \quad (6)$$

由此得到新的颜色直方图特征距离。

4.3 确定汉字区域

对聚类算法得到的属于同一类的候选区域, 提取包含所有区域的大四方区域轮廓, 形成汉字文本候选区域。

对汉字文本候选区域再次提取灰度共生矩阵, 并再次用前述的 CNN 进行分类, 判断为 H_0 类还是 H_1 类。前面 CNN 分类用到的灰度共生矩阵是由较小面积的区域计算出的, 面积小意味着计算灰度共生矩阵的元素较少, 反映了汉字纹理不一定充分, 而大区域则相对较好。因此, 用汉字文本候选区域重新提取灰度共生矩阵, 通过 CNN 再次分类来排除被错误认为是汉字文本区域的非文本区域。最终, 剩余的汉字文本候选区域就是定位到汉字文本区域。

5 实验与分析

使用本文定位方法和其他中文文本定位算法对自然场景图像中的汉字区域进行定位。图 5 给出了本文算法在一些场景图像中的定位效果图。图 5 中的汉字区域均被准确地定位出来(用黑色矩形框标出), 这主要得益于深度神经网络的分类能力, 以及 Bhattacharyya 距离对同一区域中汉字颜色的准确度量上。

实验中对比分析了本文定位算法和其他中文文本定位算法的性能。本文算法主要面向中文文



图 5 汉字区域定位结果示意图
Fig. 5 Examples for Chinese text localization

本, 目前没有统一的关于自然场景中文文本分析的标准数据库。因此, 利用照相机和网络获得 2000 幅包含各种不同类型的中文文本的自然场景图像, 包括广告牌、图书封面、路牌、商标等, 组建中文图像文本数据库(下文称作自建数据库), 人工标注自建数据库中的每幅图像每个中文文本区域的矩形框坐标。

算法性能由准确率(Precision, P)和召回率(Recall, R)两个指标反映。将定位得到矩形框 e 与标注矩形框 t 的交叉面积, 除以最小包含 e 和 t 矩形框(bounding boxes)的面积, 这个商记为 $m(e, t), 0 \leq m(e, t) \leq 1$ 。将定位算法得到的汉字区域矩形框集记为 E , 标注矩形框集为 T 。

根据下列公式确定算法的 P 和 R :

$$P = \frac{\sum_{e \in E} \max \{ m(e, t) | t \in T \}}{|E|}$$
$$R = \frac{\sum_{t \in T} \max \{ m(e, t) | e \in E \}}{|T|}$$

其中, $|E|$ 和 $|T|$ 代表矩形框集中元素的个数。
表 1 给出了本文算法和潘娜算法^[11]、徐琼算法^[13]在自建中文文本数据库上的性能对比情况。从表中可以看出, 本文算法性能优于二者, 这主要因为本文算法是基于汉字的特点提出的, 基于闭操作的部首融合和灰度共生矩阵 CNN 分类对于汉字区域定位是有效的。

表 1 自建数据集上的算法性能对比		
Tab. 1 Performance comparison between our algorithm with others on Chinese text scene image dataset		
算法	P	R
潘娜算法 ^[11]	0.76	0.72
徐琼算法 ^[13]	0.75	0.77
本文算法	0.826	0.788

实验还进行了西文字符的定位实验。数据库采用文档分析与识别国际会议(International Conference on Document Analysis and Recognition,

ICDAR) 2011^[18]。实验结果如表 2 所示,本文算法相对于课题组先前的算法^[14]有较大提高,这也是得益于 CNN 的引入;本文算法性能优于 Neumann 算法^[5],接近(但低于)Yin 算法^[9],其原因是基于闭操作的部首融合对于汉字是有效的,但对于西文的作用有限。

表 2 ICDAR 2011 数据集上的算法性能对比
Tab.2 Performance comparison between our algorithm with others on ICDAR 2011 dataset

算法	<i>P</i>	<i>R</i>
张伟伟算法 ^[14]	0.65	0.52
Neumann 算法 ^[5]	0.73	0.64
Yin 算法 ^[9]	0.86	0.68
本文算法	0.80	0.64

6 结论

本文利用场景图像中汉字区域的连通特点和纹理特性,以 CNN 分类过滤 MSER 区域为核心,提出了汉字区域定位算法。该算法的关键前导步骤是提取 MSER 区域,其对自然场景图像中纹理清晰的汉字文本区域具有较高的准确性,但一些成像质量较差、存在模糊的文本区域,存在被 MSER 区域提取漏检的情况。因此,如何减少漏检,提高定位算法召回率将是下一步的工作重点。此外,自动判断图像类型,根据类型特点选择不同定位策略,也将有助于定位性能的提高。

参考文献 (References)

[1] Karatzas D, Shafalt F, Uchida S, et al. ICDAR 2013 robust reading competition [C]//Proceedings of 12th International Conference on Document Analysis and Recognition, 2013; 1115 – 1124.

[2] Chen X R, Yuile A L. Detecting and reading text in natural scenes [C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004; 366 – 373.

[3] Pan Y F, Hou X W, Liu C L. Text localization in natural scene images based on conditional random field [C]//Proceedings of the 12th International Conference on Document Analysis and Recognition, 2009; 6 – 10.

[4] Lee J J, Lee P H, Lee S W, et al. AdaBoost for text detection in natural scene [C]//Proceedings of the International Conference on Document Analysis and Recognition, 2011; 429 – 434.

[5] Neumann L, Matas J. Real-time scene text localization and recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012; 3538 – 3545.

[6] Neumann L, Matas J. Text localization in real-world images

using efficiently pruned exhaustive search [C]//Proceedings of the International Conference on Document Analysis and Recognition, 2011; 687 – 691.

[7] Neumann L, Matas J. A method for text localization and recognition in real-world images [C]//Proceedings of the 10th Asian Conference on Computer Vision, 2010; 770 – 783.

[8] Gracia C M, Lenc K, Mimehdi M. A head-mounted device for recognizing text in natural scenes [C]//Proceedings of the 4th International Conference on Camera-based Document Analysis and Recognition, 2011; 29 – 41.

[9] Yin X C, Yin X W, Huang K Z, et al. Robust text detection in natural scene images [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36(5): 970 – 983.

[10] 刘晓佩, 卢朝阳, 李静. 结合 WTLBP 特征和 SVM 的复杂场景文本定位方法 [J]. 西安电子科技大学学报:自然科学版, 2012, 39(4): 103 – 108.

LIU Xiaopei, LU Zhaoyang, LI Jing. Complex scene text location method based on WTLBP and SVM [J]. Journal of Xidian University: Natural Science, 2012, 39(4): 103 – 108. (in Chinese)

[11] 潘娜. 图像中的文本定位算法研究 [D]. 南京: 南京理工大学, 2013.

PAN Na. Research on text detection in images [D]. Nanjing: Nanjing University of Science and Technology, 2013. (in Chinese)

[12] 孙巧榆. 复杂背景图像的文本信息提取研究 [D]. 上海: 华东师范大学, 2012.

SUN Qiaoyu. Research of the text information extraction in images with complicated background [D]. Shanghai: East China Normal University, 2012. (in Chinese)

[13] 徐琼, 于宗良, 刘峰, 等. 基于提升树的自然场景中文文本定位算法研究 [J]. 南京邮电大学学报:自然科学版, 2013, 33(6): 76 – 82.

XU Qiong, GAN Zongliang, LIU Feng, et al. Chinese text localization method based on boosting tree in natural images [J]. Journal of Nanjing University of Posts and Telecommunications: Natural Science, 2013, 33(6): 76 – 82. (in Chinese)

[14] 张伟伟, 汤光明, 孙怡峰, 等. 一种针对汉字特点的场景图像中文文本定位算法 [J]. 信息工程大学学报, 2015, 15(6): 729 – 736.

ZHANG Weiwei, TANG Guangming, SUN Yifeng, et al. Chinese scene text localization algorithm based on the feature of characters [J]. Journal of Information Engineering University, 2015, 15(6): 729 – 736. (in Chinese)

[15] Matas J, Chum O, Urban M, et al. Robust wide baseline stereo from maximally stable extremal regions [J]. Image & Vision Computing, 2004, 22(10): 761 – 767.

[16] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [C]//Proceedings of the 25th International Conference on Neural Information Processing Systems, 2012; 1097 – 1105.

[17] Bhattacharyya A. On a measure of divergence between two statistical populations defined by their probability distributions [J]. Bulletin of the Calcutta Mathematical Society, 1943, 35: 99 – 109.

[18] Shahab A, Shafait F, Dengel A. ICDAR 2011 robust reading competition challenge 2: reading text in scene images [C]//Proceedings of International Conference on Document Analysis and Recognition, 2011; 1491 – 1496.

云平台上基于关键路径截取的有向无环图应用调度算法*

刘少伟¹,任开军²,邓科峰²,宋君强²

- (1. 国防科技大学 计算机学院, 湖南 长沙 410073;
2. 国防科技大学 海洋科学与工程研究院, 湖南 长沙 410073)

摘要:针对云平台上向有向无环图科学应用执行容易产生虚拟机资源过剩、资源使用率低及费用虚高的问题,给出一种基于关键路径截取的有向无环图应用调度算法。该算法采取关键路径截取技术,循环找出最晚完成的未分配任务,从该任务出发,在所有未分配任务构成的图中找出最大连通子图,并计算该子图的关键路径,然后将关键路径上的任务集调度到性能匹配的虚拟机上执行;同时通过任务回填技术充分利用虚拟机的空闲时间槽,提高资源使用率。实验结果表明,在云计算平台上,该算法不仅能够在规定时间内完成有向无环图科学应用,而且可以提高资源使用率,有效减少完成该应用所需整体费用。

关键词:云计算平台;关键路径;虚拟机;有向无环图;资源配置
中图分类号:TP393 **文献标志码:**A **文章编号:**1001-2486(2017)03-097-08

Directed acyclic graph application scheduling strategy based on critical path cut on cloud platform

LIU Shaowei¹, REN Kaijun², DENG Kefeng², SONG Junqiang²

- (1. College of Computer, National University of Defense Technology, Changsha 410073, China;
2. Academy of Ocean Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: To address the problems that the resource is surplus, the resource utilization rate is low and the cost is unreasonably high for virtual machines in the scientific application of DAG(directed acyclic graph), a novel DAG scientific workflow scheduling algorithm based on CPC(critical path cut) was proposed. In the algorithm, the CPC technology was adopted to circularly find the unallocated task which is finished at last; the biggest connected subgraph was found from the graph constructed by the whole unallocated tasks; the critical path of this subgraph was calculated and the task set on the critical path was scheduled to the performance-matched virtual machine to execute. Meanwhile, the isolated tasks were used to fill in the idle slots of the virtual machines, such that the resource utilization could be improved. Experimental results demonstrate that, the proposed CPC algorithm can effectively reduce the execution cost of the scientific workflows while satisfying the deadline constraint in mean time.

Key words: cloud computing platform; critical path; virtual machine; directed acyclic graph; resource allocation

在许多科学研究领域,例如高能物理学、生物信息学、大气科学等,科学计算过程往往由成千上万个子任务聚合而成,而且任务之间存在严格的依赖关系。这些子任务含有依赖关系的大规模科学应用可以抽象为大规模有向无环图(Directed Acyclic Graph, DAG)科学应用。

这些应用,如天文学应用 Montage^[1]、天体物理学应用激光干涉引力波观察台(Laser Interferometer Gravitational wave Observatory, LIGO)^[1],通常需要在复杂的分布式计算机系统上执行,例如超级计算机、分布式集群系统以及网格系统等。针对这种应用,目前已经有很多成熟的算法,例如文献[2]提出了一种服务计数算法(Server Count Bound, SCB)启发式算法,在服从截止时间约束的同时,寻找最小资源分配方案以减少用户费用,可以在集群上最小化资源使用。文献[3]在费用约束下通过数据还原及重新利用技术提高 DAG 的执行效率,侧重于对 DAG 中节点的重复利用以减少计算开销。文献[4]则主要针对 DAG 应用,通过分析应用的同步完成特征,提出了三种针对性算法。

但是,上述算法主要运行平台为高性能集群或超级计算机,构造这样的高性能计算平台往往代价异常昂贵,对其访问一般也需要复杂耗时的

* 收稿日期:2016-02-14
基金项目:国家自然科学基金资助项目(61572510);国家公益行业专项计划资助项目(GYHY201306003)
作者简介:刘少伟(1987—),男,陕西渭南人,博士研究生,E-mail:liushaowei@nudt.edu.cn;
宋君强(通信作者),男,研究员,硕士,博士生导师,E-mail: junqiang@nudt.edu.cn

申请过程。

近年来,云计算的兴起为大规模 DAG 科学应用的高性价比执行提供了潜力。云计算技术是一种共享基础架构的方法,它通过虚拟化技术将计算资源和存储资源虚拟成一个资源池,以虚拟机(Virtual Machine, VM)的形式向用户提供资源。这种资源提供方式降低了供应商运行成本,同时用户可以根据自己需求合理配置自己所需要的资源。因此,云计算吸引着越来越多的用户将大规模 DAG 科学应用迁移到云平台上执行。但同时,云计算的资源供应模式和收费模式为大规模 DAG 科学应用的运行带来了新的挑战。

Byun 等针对云计算环境提出了分层的负载均衡时间调度(Partitioned Balanced Time Scheduling, PBTS)算法^[5],PBTS 算法将 workflow 截止期划分为多个时间段,并利用 BTS 算法^[6]为每个时间段计算 DAG 应用需要的最少资源量,但算法针对的是同构资源模型。针对云计算平台虚拟机资源的获取与任务调度完全由用户负责的特点,文献[7]提出了一种云平台上基于截止时间分发的偏序关键路径调度(IaaS Cloud Partial Critical Paths with Deadline Distribution, IC-PCPD2)算法,首先初始化关键路径上任务的截止时间,然后采用递归的方法依次求取其他路径上任务的截止时间,然后循环求取偏序关键路径 PCP,在截止时间约束下将 PCP 放置到最便宜的虚拟机实例上,直到 DAG 中所有任务分配完毕。由于是按路径分配任务,该算法容易造成更多的虚拟机空闲碎片,降低资源使用率,相对地增加了用户费用。

本文以最小化完成 DAG 所需费用为目标,通过分析大规模 DAG 科学应用中任务的依赖关系,提出一种云平台上基于关键路径截取(Critical Path Cut, CPC)的 DAG 调度算法。

1 模型定义及问题分析

1.1 相关模型

1.1.1 应用模型

针对大规模 DAG 科学应用,使用符号 W 进行指代, $W = \{V, E\}$ 。其中, $V = \{t_i | i = 1, 2, \dots, |V|\}$ 是顶点的集合,表示 DAG 应用中任务集合; $E = \{e_k | k = 1, 2, \dots, |E|\}$ 为边的集合,表示任务间的控制或数据依赖关系。单个任务进一步由五元组描述: $t_i = \{seq, \omega, \lambda, EST, LST\}$ 。其中, $seq(t_i)$ 表示任务的串行执行时间; $\omega(t_i)$ 表示任务

的实际执行时间; $\lambda \in [0, 1]$ 为并行系数, $\omega(t_i) = seq(t_i) \cdot \lambda$; $EST(t_i)$ 表示任务的最早开始时间; $LST(t_i)$ 表示任务的最晚开始时间。

边表示为 $e_k = (t_i, t_j)$, 其中 t_i 为 e_k 的起点, t_j 为 e_k 的终点, t_j 只有在 t_i 执行完毕之后才可以开始执行; $\omega(e_k)$ 为权重,表示两个任务之间的通信开销,当任务 t_i 和 t_j 调度到同一个虚拟机上时,通信开销忽略不计。

无任何前驱的任务为入口任务,无任何后继的任务为出口任务。入口任务和出口任务具有唯一性,如果 DAG 包括多个入口任务或出口任务,可以将所有入口任务和出口任务连接到一个零成本的虚拟入口任务 t_{start} 或虚拟出口任务 t_{finish} ,对系统性能并无影响。

1.1.2 资源模型

假设云计算平台提供 N 种不同类型的虚拟机,表示为 $vmType_i = \{cpuNum, memory, storage, price\}$, $i = 1, 2, \dots, N$ 。不同类型的 VM, 参数值各不相同,其中 $cpuNum$ 表示 CPU 数量, $memory$ 表示内存大小, $storage$ 表示硬盘大小, $price$ 表示此种类型的虚拟机单个时间周期费用。

用户通过租赁云计算平台的虚拟机构建虚拟机集群执行科学 workflow, 虚拟机集群表示为 $VMC = \{vm_i | i = 1, 2, \dots\}$, $vm_i = \{id, vmType, sTime, aTime, rTime, scheTL\}$, 其中 id 表示虚拟机编号, $vmType$ 表示虚拟机类型, $sTime$ 表示虚拟机启动时间, $aTime$ 表示虚拟机生存时间, $rTime$ 表示虚拟机当前时间周期的剩余时间, $scheTL$ 表示调度到虚拟机的任务集合。

考虑现代虚拟化技术下虚拟机的启动时间和关闭时间已经达到秒级,因此假设虚拟机启动时间和关闭时间可以忽略不计。

1.2 符号定义

本节以图 1 的简单 DAG 为示例,对本文用到的参数、符号进行说明。

图中 $t_1 \sim t_9$ 为 DAG 中的任务集合, t_{start} 和 t_{finish} 为新增的开始任务和结束任务,箭头表示任务间依赖关系。为了便于说明和理解,本示例中将通信时间略去。假设云平台上有两种类型的虚拟机,一种只有 1 个 VCPU,另一种类型虚拟机拥有 2 个 VCPU。如图 1 所示,括号内、外分别是任务在双核和单核虚拟机实例上的运行时间,单位为 min。虚拟机实例需要按小时申请。假设整个 DAG 的截止时间(deadline)为 120 min。

1.2.1 任务并行性

任务并行性包括三个方面:多任务并行、单任

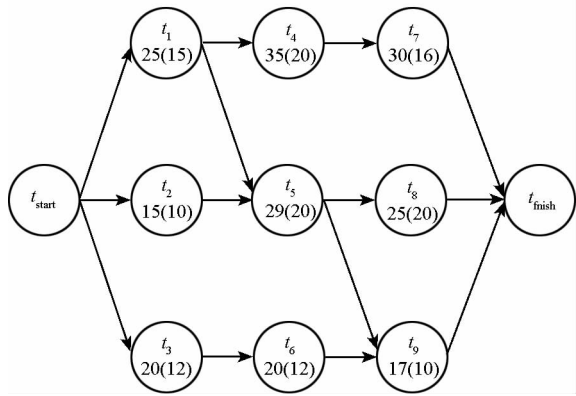


图1 简单 DAG 示例

Fig. 1 Simple DAG example

务多核并行和单任务多虚拟机并行。

多任务并行指不同任务在多个虚拟机上并发执行,如图1中 t_1 和 t_2 可以在不同虚拟机上并发执行;单任务多核并行指单个任务可以在拥有多个VCPU上的虚拟机上并行执行,如图1中 t_1 在单核虚拟机上执行时间为25 min,在双核虚拟机上执行时间为15 min;单任务多虚拟机并行指单个任务可以在多个虚拟机上通过并行执行减少任务执行时间。

1.2.2 关键路径

将 DAG 图中从开始任务 t_{start} 到结束任务 t_{finish} 所需时间最长的一条路径定义为关键路径。DAG 应用的关键路径决定了其最短执行时间。计算关键路径首先需要对所有任务的 EST 和 LST 进行赋值,其中最早开始时间 EST 计算方式如下:

$$EST(t_i) = \begin{cases} 0, & t_i = t_{start} \\ \max\{EST(t_j) + \omega(t_j) + \omega(e_k)\}, & e_k = (t_j, t_i) \in E \end{cases}$$

最晚开始时间 LST 计算方式如下:

$$LST(t_i) = \begin{cases} EST(t_i), & t_i = t_{finish} \\ \min\{LST(t_j) - \omega(t_j) - \omega(e_k)\}, & e_k = (t_j, t_i) \in E \end{cases}$$

关键路径计算方式分以下6步:①计算所有任务的 EST ;②计算所有任务的 LST ;③计算所有边的 EST ;④计算所有边的 LST ;⑤所有 LST 等于 EST 的边构成关键路径。

需要注意的是,如果任务的实际执行时间不固定,那么关键路径也可能发生变化。如图1所示 DAG,如果所有任务都在单核虚拟机上执行,那么关键路径为 $\{t_{start}, t_1, t_4, t_7, t_{finish}\}$;如果所有任务都在双核虚拟机上执行,关键路径则为 $\{t_{start}, t_1, t_5, t_8, t_{finish}\}$ 。

1.2.3 截止时间

用户一般会指定一个截止时间 δ ,其取值需要在合理范围之内,本文采取以下取值方法。

假设所有任务的实际执行时间 $\omega(t_i)$ 为其在配置最高的虚拟机实例上的执行时间,此时完成关键路径上任务所需时间为 δ_{short} ,即最快完成时间;假设所有任务的实际执行时间 $\omega(t_i)$ 为其在配置最低的虚拟机实例上的执行时间,此时完成关键路径上任务所需时间为 δ_{long} ,即最慢完成时间。

可以看出,如果 $\delta < \delta_{short}$,则无论如何调度,DAG 应用都无法在截止时间 δ 约束下完成;如果 $\delta_{short} \leq \delta \leq \delta_{long}$,为了保证按时完成 DAG 应用,肯定需要配置高性能的虚拟机实例;如果 $\delta > \delta_{long}$,所生成的虚拟机集群一般都是最低配置的虚拟机实例,以实现费用最小化。

1.3 问题分析

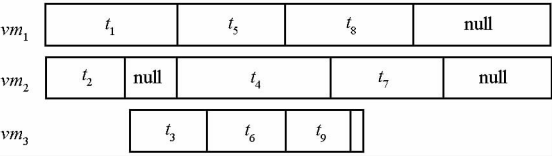
DAG 应用在云计算平台执行时,调度算法需要解决以下几个方面所面临的问题:

1)配置虚拟机集群。既不能启动大量 VM,导致相当一部分 VM 空转,造成资源浪费,也不能启动少量 VM,导致无法按时完成 DAG 科学应用。调度算法必须合理规划好每个 VM 的类型、启动时间和生存时间,在满足整个 DAG 应用截止时间约束的情况下最大化每个虚拟机的成本效益。

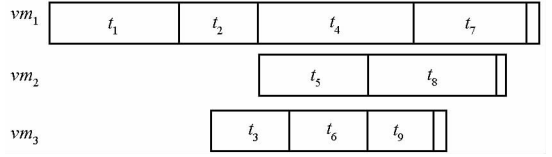
2)任务调度。调度父任务到低性能虚拟机上时,可能引起后续子任务长时间等待,导致 DAG 应用无法在截止时间 δ 内完成;调度任务到高性能虚拟机上执行,由于任务很快执行完毕,当前虚拟机因为等待其他虚拟机上的父任务完成而处于空转状态,则会造成资源浪费,间接增加用户费用。

3)减少空闲时间槽,提高资源使用率。当前大多数云计算平台采用基于时间周期计费的方式,如 Amazon EC2 虚拟机实例以小时为单位计费,不足1小时按1小时计算。在此情况下,调度算法必须有效地将任务整合在一起,使得租赁的虚拟机能够在整数时间周期内充分利用。

4)以图1所给 DAG 为例。IC-PCPD2 算法调度方案如图2(a)所示。IC-PCPD2 总共需要5(cpu·hour)的资源,可以看出 IC-PCPD2 算法会产生部分虚拟机碎片。实际上如果按照图2(b)所示的新方案调度,只需要4(cpu·hour)的资源,而且能够在截止时间内完成 DAG 应用所有任务,费用可以节省20%。因此,为了进一步提高虚拟机空闲时间槽的利用,减少执行开销,本文提出了基于关键路径截取的 DAG 调度算法,即 CPC 算法。



(a) IC-PCPD2 算法
(a) IC-PCPD2 algorithm



(b) 新解决方案
(b) New solution

图 2 两种调度方案

Fig. 2 Two kinds of scheduling schemes

2 算法描述

CPC 算法的目标是在截止时间条件约束下,以费用最小化为目标,完成 DAG 应用。

CPC 算法吸取了 IC-PCPD2 路径调度的思想,但略有不同。在路径调度方法中,首先找出最迟完成的未分配任务,并从该任务出发,找出未分配任务构成的最大连通子图,求取该子图的关键路径,并进行合理调度;同时增加了任务回填方法,通过任务回填技术将单任务填充到已有虚拟机空闲时间槽,提高对虚拟机资源的使用率。

2.1 CPC 调度算法

算法 1 为基于 CPC 的 DAG 调度算法。其中算法输入为 DAG 应用节点 V 和边 E 的情况、所有边的权重 $\omega(e_k)$ 、用户给出的截止时间 δ ;算法输出为 CPC 算法生成的虚拟机集群及任务到具体虚拟机的调度方案。

算法主要分以下几步。第一步(语句 1 ~ 13):在截止时间约束下,通过计算关键路径长度确定第一个虚拟机实例 vm_1 的类型;然后将关键路径上的任务调度到 vm_1 上,同时更新 vm_1 各项参数,并将 vm_1 加入到 $vmList$ 中。其中,语句 2 ~ 4 计算关键路径总权重,语句 5 ~ 12 设置 vm_1 的相关信息。第二步(语句 14):将未分配的任务加入到 $remainTaskList$ 列表。第三步(语句 15 ~ 23):首先采取任务回填方法,尝试将未分配的单个任务调度到已存在的 $vmList$ 中;然后使用路径调度方法,在剩余未分配的任务构成的子图中,计算出一条关键路径,新增虚拟机实例来完成任务。

算法 1 CPC 算法

Alg. 1 CPC algorithm

Name: CPC; Critical path cut algorithm
Input: $W = \{V, E\}$; every $\omega(e_k)$; δ (deadline)
Output: $vmList = \{vm_k \mid k = 1, 2, \dots\}$; $vm_k = \{id, vmType, sTime, aTime, rTime, scheTL\}$

```
1  FOR (from  $vmType_1$  to  $vmType_N$ )
2    FOR(each  $t_i \in V$ )  $\omega(t_i) = seq(t_i) * \lambda(vmType_i)$ 
3     $\{kpEL, kpTL\} = getCriticalPath(V, E)$ 
4     $exeTime = \sum_{t_i \in kpTL} \omega(t_i) + \sum_{e_k \in kpEL} \omega(e_k)$ 
5    IF ( $exeTime < \delta$ )
6      New first VM:  $vm_1$ ;
7       $vm_1.vmType = curType$ ; //vm type of first VM
8       $vm_1.scheTL = kpTL$ ; //task scheduled on  $vm_1$ 
9      update  $vm_1$ 's  $sTime, aTime$  and  $rTime$ ;
10     add  $vm_1$  to  $vmList$ ;
11     break;
12   END IF
13 END FOR
14 update  $remainTaskList$ ;
15 While(exist( $t_i$ ) in  $remainTaskList$ )
16   FOR(each  $t_i \in remainTaskList$ )
17     FOR(each  $vm_k \in vmList$ )
18       placeTaskOnExistVM(); //schedule single task
19     END FOR
20   END FOR
21   placeTaskOnNewVM(); //schedule path
22   update  $remainTaskList$ ;
23 END While
```

2.2 任务回填算法

CPC 算法第 18 步算法 placeTaskOnExistVM 尝试将单个任务放到已经申请的虚拟机空闲时间槽中,以提高资源利用率,如算法 2 所示。

该算法尝试将 t_i 调度到 vm_k 上。语句 1 计算 t_i 在 vm_k 上的实际执行时间。语句 3 ~ 8 对 t_i 在 vm_k 的 $number(vm_k.scheTL) + 1$ 个可能的插入位置进行逐一尝试。

其中 $canPlace(t_i, j)$ 计算是否能够将 t_i 放在 vm_k 的任务列表 $scheTL$ 的第 j 个位置,并返回一个布尔值。特别地,只有当以下两个条件同时满足时, $canPlace(t_i, j)$ 才返回 true,调度成功:① t_i 插入到 $scheTL$ 的第 j 个位置后, vm_k 上 $scheTL$ 中位于 t_i 前面的任务不包括 t_i 的子任务,位于 t_i 后面的任务不包括 t_i 的父任务,保证任务间存在的依赖关系;② t_i 插入到 $scheTL$ 的第 j 个位置时,计算所有任务的 EST 和 LST ,保证 $\max\{LST(t_i) +$

$\omega(t_i)$ 的值小于截止时间 δ 。

算法 2 任务回填算法

Alg.2 Task backfill algorithm

```
Name: placeTaskOnExistVM
Input:  $t_i$ ;  $vm_k$ ;  $W = (V, E)$ 
Output: flag;

1  $\omega(t_i) = seq(t_i) * \lambda(vm_k, vmType)$ ;
2 IF ( $vm_k.rTime > \omega(t_i)$ )
3   FOR ( $j = 0$ ;  $j \leq number(vm_k.scheTL)$ ;  $j++$ )
4     IF ( $canPlace(t_i, j)$ )
5       Insert( $vm_k.scheTL, j, t_i$ );
6       return true;
7   END IF
8 END FOR
9 return false;
10 ELSE return false;
```

2.3 路径调度算法

当单个任务无法回填到已有虚拟机的空闲时间槽时, CPC 算法启动路径调度算法 placeTaskOnNewVM。该方法在剩余未分配的任务构成的子图中, 计算出一条关键路径, 并新增虚拟机实例来完成这条关键路径上的任务, 如算法 3 所示。

算法 3 路径调度算法

Alg.3 Path scheduling algorithm

```
Name: placeTaskOnNewVM
Input: remainTaskList; vmList;  $W = (V, E)$ ;
 $\delta$  (deadline);
Output: empty

1 FOR (from  $vmType_1$  to  $vmType_N$ )
2   FOR (each  $t_i \in remainTaskList$ )  $\omega(t_i) = seq(t_i) * \lambda(currentVmType)$ ;
3   calculate all tasks' EST;
4   calculate all tasks' LST
5   FOR (each  $t_i \in remainTaskList$ )
6     FinishTime( $t_i$ ) = LST( $t_i$ ) +  $\omega(t_i)$ ;
7     IF (FinishTime( $t_{max}$ ) < FinishTime( $t_i$ ))  $t_{max} = t_i$ ;
8   END FOR
9   IF (FinishTime( $t_{max}$ ) <  $\delta$ )
10    childGraph = getChildGraph( $t_{max}, remainTaskList, E$ );
11    childCriticalPath = getCriticalPath(childGraph);
12    new VM:  $vm_k$ ;
13     $vm_k.vmType = curType$ ; //vm type
14    update  $vm_k$ 's sTime, aTime, rTime
15     $vm_k.scheTL = childCriticalPath$ ; //task
    scheduled on  $vm_k$ 
16    add  $vm_k$  to vmList;
17    break;
18   END IF
19 END FOR
```

算法中语句 2 ~ 4 计算出所有任务的 *EST* 和 *LST* 值; 语句 5 ~ 8 找出 *remainTaskList* 完成时间最晚的任务 t_{max} , $FinishTime(t_i)$ 表示任务执行完毕的时间点; 语句 10, 从 t_{max} 出发, 在 *remainTaskList* 中找到最大连通子图; 语句 11, 计算该连通子图的关键路径, 同时计算出该关键路径的开始时间和结束时间; 语句 12 ~ 16, 启动一个新的 VM 以执行连通子图关键路径上的任务, 并将该 VM 加入 *vmList*。

2.4 简单示例

以图 1 所给 DAG 为例, CPC 算法具体执行过程如图 3 所示。

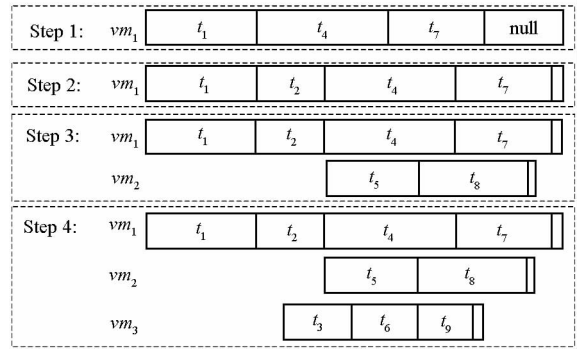


图 3 CPC 算法示例

Fig.3 Example for CPC algorithm

第一步: 找到 DAG 中最晚完成的任务 t_7 , 并从 t_7 出发找到最大连通子图, 求得关键路径 $\{t_{start}, t_1, t_4, t_7, t_{finish}\}$, 然后根据关键路径长度申请虚拟机 vm_1 , 虚拟机类型为单核, 生存时间为 120 min。第二步: 尝试将其余任务放入 vm_1 , 发现 t_2 符合条件, 放置成功。第三步: 从未分配的任务中找到最晚完成的任务 t_8 , 然后从 t_8 出发找到最大连通子图, 并求得关键路径 $\{t_5, t_8\}$, 再根据关键路径长度申请虚拟机 vm_2 , 虚拟机类型为单核, 生存时间为 60 min; 尝试将其余任务放置到 vm_2 , 剩余空间不够, 适配失败。第四步: 找到未分配任务中最晚完成的任务 t_9 , 并从 t_9 出发找到最大连通子图, 求得关键路径 $\{t_3, t_6, t_9\}$, 然后根据关键路径长度申请虚拟机 vm_3 , 虚拟机类型为单核, 生存时间为 60 min; 任务分配完毕, 算法结束。

可以看出, CPC 算法能尽量减少虚拟机的空闲时间碎片, 最可能提高资源使用率, 减少用户使用费用。

2.5 时间复杂度

如算法 1 所示, CPC 算法中语句 1 对可选的虚拟机类型进行遍历, 由于可选类型一般不超过 10, 所以复杂度为 $O(1)$; 语句 3 计算关键路径, 复杂度

为 $O(n^2)$, 因此语句 1 ~ 13 复杂度为 $O(n^2)$ 。第 18 句任务回填中需要计算所有任务的 EST 和 LST , 计算复杂度为 $O(q \cdot n)$, 其中 q 是每个 VM 上任务的平均数量; 第 21 句任务调度需要计算剩余任务的关键路径, 复杂度为 $O(k^2)$; 那么 15 ~ 23 句的复杂度为 $O(k) \cdot O(p) \cdot O(q \cdot n) + O(k^2) = O(k \cdot p \cdot q \cdot n) + O(k^2)$, 其中 p 为已存在的 VM 数量, k 为剩余任务数量, 那么有 $k = n - p \cdot q$, 所以 $O(k \cdot p \cdot q \cdot n) + O(k^2) = O(k \cdot (n - k) \cdot n) + O(k^2) = O(n^2k - nk^2) + O(k^2)$, 其中 k 平均值为 $n/2$, 所以 15 ~ 23 句复杂度为 $O(n^3)$ 。

综上, CPC 算法整体时间复杂度为 $O(n^3)$ 。

3 算法验证

3.1 测试用例

采用真实的 DAG 科学应用 Montage 科学工作流来对算法进行验证。Montage 科学工作流是 NASA/IPAC 开发的一个天文学应用, 该应用能够利用 FITS 格式的图像生成自定义的天空拼接图。

3.2 云平台仿真

采用云仿真工具 CloudSim^[8] 对云计算平台进行模拟, 主要使用 VirtualMachine 类模拟虚拟机, 使用 VMCharacteristics 类对虚拟机特征进行描述, 并在 VMScheduler 类中实现不同的调度策略。

计费模式采用 Amazon EC2 的计费模式, 即虚拟机实例按时间周期计费, 单 VCPU 虚拟机实例当前价格取 0.16 \$/h。实验使用的四种通用类型虚拟机如表 1 所示。

表 1 费用模式
Tab.1 Cost model

VM 类型	性能指标	价格/ (\$ /h)
type 1	1 VCPU, 1 G 内存, 50 G 硬盘空间	0.16
type 2	2 VCPU, 2 G 内存, 100 G 硬盘空间	0.32
type 3	4 VCPU, 8 G 内存, 200 G 硬盘空间	0.64
type 4	8 VCPU, 16 G 内存, 500 G 硬盘空间	1.28

3.3 对比算法

采用混合最早完成时间 (Heterogeneous Earliest Finish Time, HEFT), IC-PCP 和 IC-PCPD2 算法作为对比算法。

HEFT 算法: 根据任务的平均执行时间和通信时间计算出一个 HEFT 值, 在任务分配时, 选择具有最大 HEFT 值的任务并将其调度到完成时间最小的虚拟机上。

IC-PCP 和 IC-PCPD2 算法: 首先初始化关键路径上任务的截止时间, 然后采用递归的方法依次求取其他任务的截止时间, 之后依次求取偏序关键路径 PCP, 在截止时间约束下将 PCP 路径上任务放置到可将其完成的最便宜的虚拟机上执行, 其中 IC-PCP 算法每次尝试将该路径上的任务放置到虚拟机实例上, 而 IC-PCPD2 算法则是计算偏序关键路径上的任务的子截止时间, 并对任务的子截止时间进行设定, 任务的分配则根据其子截止时间进行合理调度。IC-PCP 和 IC-PCPD2 算法复杂度为 $O(n^2)$ 。

3.4 实验结果

本文主要对比两个实验参数: 总体费用和资源使用率。

总体费用指完成同一个科学工作流所需的整体费用, 即 $费用 = \sum_{i=1}^N price_{vm_i} \cdot aTime_{vm_i}$, 其中 $price_{vm_i}$ 表示 vm_i 单位时间的收费标准, $aTime_{vm_i}$ 表示 vm_i 的生存时间。资源使用率指 CPU 的实际使用占所申请的整体 CPU 资源的比例, 即:

$$资源使用率 = 1 - \frac{\sum_{j=1}^N rTime_{vm_j} \cdot cpuNum_{vm_j}}{\sum_{i=1}^N aTime_{vm_i} \cdot cpuNum_{vm_i}}$$

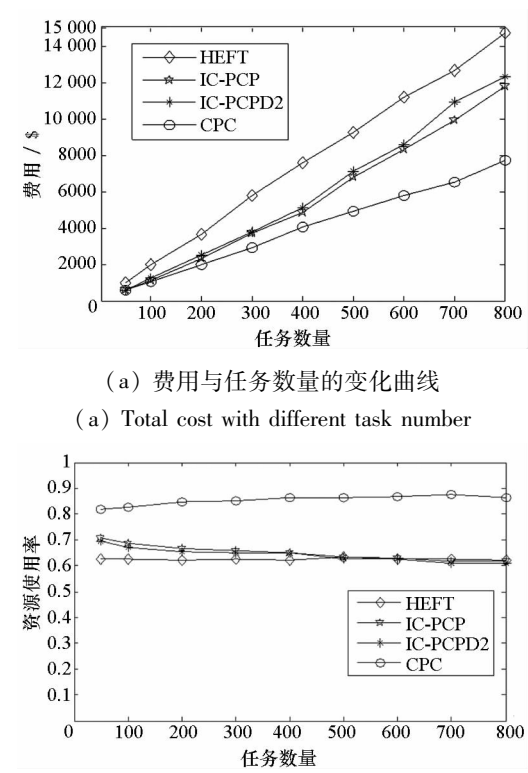
式中: $rTime_{vm_j}$ 表示 vm_j 的剩余时间, 即空闲时间。

3.4.1 任务数量变化对结果的影响

实验中, 虚拟机实例类型采用表 1 所给, 虚拟机实例时间周期长度为 60 min, 截止时间取 $\delta \in [\delta_{short}, \delta_{long}]$, 任务数量取值范围为 {50, 100, 200, 300, 400, 500, 600, 700, 800}, 任务并行比例 η 服从 [0.1, 0.9] 的均匀分布。实验结果如图 4 所示。

从图 4(a) 可以看出, 随着 Montage 任务数量的增加, HEFT, IC-PCP, IC-PCPD2 和 CPC 算法完成 DAG 应用所需费用几乎线性增加。由于时间周期为较长的 60 min, IC-PCP 略优, 比 IC-PCPD2 费用减少约 2%; 而 CPC 算法比 IC-PCP 算法费用减少了约 3% ~ 29%。从图 4(b) 可以看出, CPC 算法的资源使用率在 80% 左右, 而 IC-PCP 和 IC-PCPD2 算法的资源使用率为 57% ~ 70%。

这是因为 CPC 算法不仅采用关键路径调度, 对大堆任务进行合理调度, 同时还采用灵活的任务回填方法, 尝试将单个任务放置到虚拟机时间空闲槽, 成功率高。因此 CPC 算法对虚拟机实例的空闲时间槽利用率高, 从而提高了资源使用率。而且任务数量越大, 所需虚拟机实例越多, 这种优势就越明显。



(a) 费用与任务数量的变化曲线
(a) Total cost with different task number

(b) 资源使用率与任务数量的变化曲线
(b) Resource utilization with different task number

图4 任务数量变化对结果影响

Fig.4 Results with different task number

3.4.2 DAG 截止时间变化对结果的影响

实验中,虚拟机实例类型采用表1所给,任务数量选取200,任务并行比例 η 服从 $[0.1,0.9]$ 的均匀分布,虚拟机实例时间周期长度为60 min。实验结果如图5所示,图中, $S = \delta_{\text{short}}$, $d = (\delta_{\text{long}} - \delta_{\text{short}})/9$ 表示将 $[\delta_{\text{short}}, \delta_{\text{long}}]$ 划分为9等份,那么 $\delta_{\text{long}} = S + 9d$ 。

从图5(a)可以看出,随着截止时间 δ 取值的增加,HEFT所需费用越小,而IC-PCP, IC-PCPD2和CPC算法变化不大,这是因为后三种算法已经最大化利用了费用优化空间。图5(b)反映了CPC算法的资源使用率相比IC-PCP、IC-PCPD2和HEFT算法的较优。

这是因为IC-PCP、IC-PCPD2和CPC算法都采用了路径调度策略,分别计算了偏序关键路径和关键路径,完工时间与路径长度相关,充分利用了优化空间,受截止时间影响较小。即使截止时间变长,对费用和资源使用率的影响也有限。

3.4.3 时间周期变化对结果的影响

实验中,虚拟机实例类型采用表1所给,截止时间 $\delta \in [\delta_{\text{short}}, \delta_{\text{long}}]$,任务数量选取200,任务并行比例 η 服从 $[0.1,0.9]$ 的均匀分布,虚拟机实例时间周期长度选取 $\{60, 50, 40, 30, 20, 10, 5\}$,

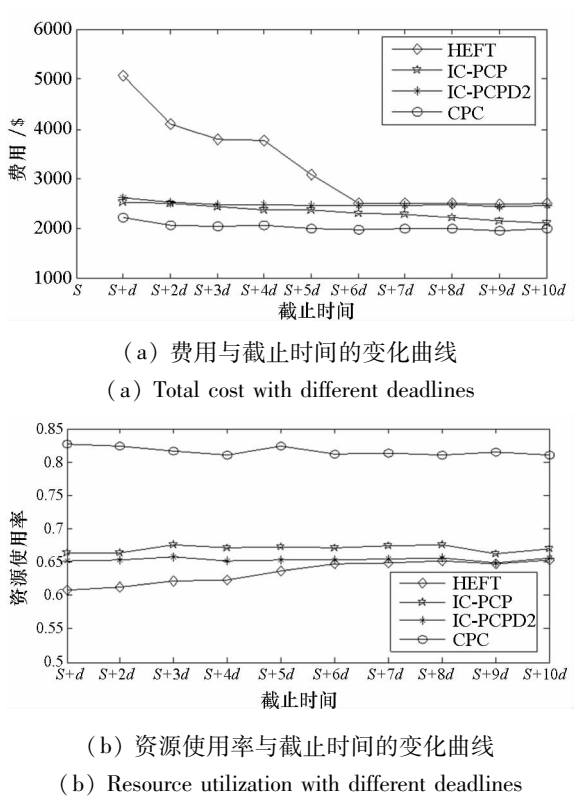
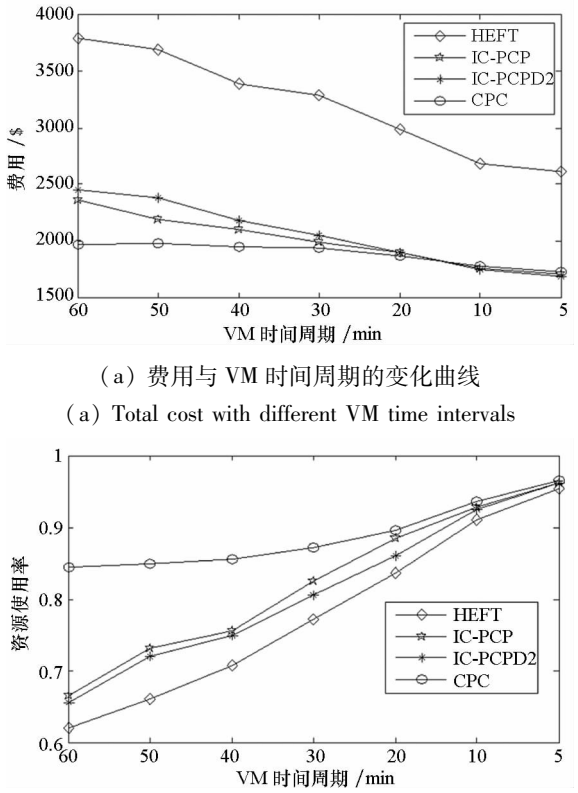


图5 截止时间 δ 对结果的影响
Fig.5 Results with different deadline δ

单位为 min。实验结果如图6所示。



(a) 费用与 VM 时间周期的变化曲线
(a) Total cost with different VM time intervals

(b) 资源使用率与 VM 时间周期的变化曲线
(b) Resource utilization with different VM time intervals

图6 VM 时间周期长度对结果的影响

Fig.6 Results with different time intervals of VM

从图 6(a)可以看出,随着虚拟机时间周期长度的减少,HEFT、IC-PCP、IC-PCPD2 和 CPC 算法的费用也不断减少。时间周期为 30 ~ 60 min 时,CPC 算法比 IC-PCP 算法所需费用少约 4% ~ 13%,但时间周期为 5 ~ 20 min 时,两者费用相当。从图 6(b)看出,四种算法在 VM 时间周期减少的过程中,资源使用率不断提高;时间周期越短,CPC 算法的优势就越小。

这是因为时间周期越短,虚拟机实际空转时间变短,资源利用率就越高,因此所需的花费越少。

4 结论

本文研究云计算平台上截止时间约束下 DAG 科学应用的费用优化问题,根据云计算平台资源特性,设计了基于关键路径截取的 DAG 调度算法。该算法通过分析任务间依赖关系,采取关键路径截取技术将任务聚合在匹配的虚拟机上执行,通过路径调度完成任务集到资源的映射,同时通过任务回填方法将单个任务调度到已分配虚拟机的空闲时间槽上执行,提高了虚拟机的资源利用率。最后,利用真实科学工作流 Montage 对 CPC 算法进行了验证。结果表明,CPC 算法可以根据 DAG 应用内在特点,配置合适的虚拟机集群资源,并对任务到虚拟机进行合理调度,不仅能够在规定时间内完成 DAG 应用,而且可以有效减少执行费用。

参考文献 (References)

[1] Bharathi S, Chervenak A, Deelman E, et al. Characterization

of scientific workflows[C]//Proceedings of Third Workshop on Workflows in Support of Large-Scale Science, 2008: 1 – 10.

[2] Moens H, Handekyn K, De Turck F. Cost-aware scheduling of deadline-constrained task workflows in public cloud environments [C]//Proceedings of IFIP/IEEE International Symposium on Integrated Network Management (IM 2013), 2013: 68 – 75.

[3] Ma Y L, Shi M Y, Wei J. Cost and accuracy aware scientific workflow retrieval based on distance measure[J]. Information Sciences, 2015, 314: 1 – 13.

[4] 刘灿灿, 张卫民, 骆志刚. 基于逆向分层的工作流时间 – 费用优化方法[J]. 国防科技大学学报, 2013, 35(3): 61 – 66.

LIU Cancan, ZHANG Weimin, LUO Zhigang. Time and cost trade-off heuristics for workflow scheduling based on bottom level[J]. Journal of National University of Defense Technology, 2013, 35(3): 61 – 66. (in Chinese)

[5] Byun E K, Kee Y S, Kim J S, et al. Cost optimized provisioning of elastic resources for application workflows[J]. Future Generation Computer Systems, 2011, 27(8): 1011 – 1026.

[6] Byun E K, Kee Y S, Kim J S, et al. BTS: resource capacity estimate for time-targeted science workflows[J]. Journal of Parallel and Distributed Computing, 2011, 71(6): 848 – 862.

[7] Abrishami S, Naghibzadeh M, Epema D H J. Deadline-constrained workflow scheduling algorithms for infrastructure as a service clouds[J]. Future Generation Computer Systems, 2013, 29(1): 158 – 169.

[8] Calheiros R N, Ranjan R, Beloglazov A, et al. CloudSim: a toolkit for modeling and simulation of cloud computing environments and evaluation of resource provisioning algorithms [J]. Software: Practice and Experience, 2011, 41(1): 23 – 50.

滑模控制的新型双幂次组合函数趋近律*

廖 瑛¹,杨雅君^{1,2},王 勇¹

(1. 国防科技大学 航天科学与工程学院, 湖南 长沙 410073; 2. 装备学院, 北京 101416)

摘 要:提出一种基于双幂次组合函数趋近律的新型滑模控制方案。与现有的快速幂次或双幂次趋近律相比,具有更快的收敛速度,同时还保持了全局固定时间收敛特性,收敛时间上界与滑模初值无关。当系统存在有界扰动时,能够使滑模变量在有限时间内收敛到稳态误差界内,同时其稳态误差要小于现有方法的。仿真实验验证了该方法的有效性 & 理论分析的正确性。

关键词:幂次趋近律;双幂次趋近律;固定时间收敛;稳态误差界;非线性组合函数
中图分类号:TP273 **文献标志码:**A **文章编号:**1001-2486(2017)03-105-06

Novel double power combination function reaching law for sliding mode control

LIAO Ying¹, YANG Yajun^{1,2}, WANG Yong¹

(1. College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China;
2. Equipment Academy, Beijing 101416, China)

Abstract: A novel sliding control approach based on double power combination function reaching law was proposed. The proposed reaching law has faster convergence speed in comparison with fast power/double power reaching law, and it also has the characteristic of global fixed-time convergence, which means the upper bound of convergence time is independent of the initial value of sliding mode variables. It was proved that for a class of bounded external disturbance, the sliding mode variable can converge to a proposed steady-state error bounds in finite time, and the value of the steady-state error is less than the previous reaching law. Simulation results show that the validity of conclusion is confirmed.

Key words: power rate reaching law; double power rate reaching law; fixed time convergence; steady-state error bound; nonlinear combination function

滑模控制是一种鲁棒控制方法,在滑模运动阶段对系统中的匹配扰动项具有不变性,广泛应用于不确定性系统的控制问题。传统滑模控制在系统状态处于滑模面上时,产生高频切换的控制信号,在保证滑动模态存在的同时也引发了严重的抖振现象。对实际控制对象而言,抖振不仅意味着过高的能量消耗,也容易激发未建模的高频动态而导致系统失稳。此外,控制的鲁棒性只在滑模运动阶段存在,在滑模趋近阶段系统仍受到不确定性和外扰的影响。如何缩短趋近阶段的时间和消除抖振,一直是滑模控制研究的热点问题。目前解决该问题常见的方法有:准滑模(边界层)方法^[1]、高阶滑模控制^[2]、非奇异终端滑模控制^[3]、动态滑模控制^[4]和趋近律技术^[5]。准滑模方法利用饱和函数或连续函数近似传统滑模控制中的符号函数,使系统状态进入并保持在滑模

面周围的邻域内,即形成所谓“准滑模”运动,有效削弱了抖振,但系统只能达到一致有界稳定,事实上降低了控制精度。高阶滑模控制和动态滑模控制将产生控制切换信号的符号函数置于控制输入的一阶或更高阶导数上,避免了抖振现象,但难以应用于一阶系统,获取滑模变量高阶导数信号也存在一定难度。非奇异终端滑模控制既能有效去除抖振也能够在规定时间内使系统状态收敛于平衡点,但相对传统滑模的指数趋近律,其收敛速度非常慢,实际上降低了滑模趋近阶段的过渡品质。高为炳^[5]提出了趋近律技术的概念并分析了等速趋近律、指数趋近律和幂次趋近律等方法。其中:等速趋近律可视为传统滑模控制,趋近速度恒定,指数趋近律通过增加线性项加快了状态远离滑模面时的趋近速度,这两种方法均不能完全消除抖振;幂次趋近律中符号函数的增益与滑模

* 收稿日期:2016-02-17
基金项目:航天科技创新基金资助项目(CAST201502)
作者简介:廖瑛(1961—),女,湖南长沙人,教授,博士,博士生导师,E-mail:liaoqing1104@163.com

变量绝对值的幂次成正比,在状态到达滑模面时趋近速度为零,消除了抖振,但在远离滑模面时趋近速度较小。结合指数趋近律与幂次趋近律,文献[6]提出了一种快速幂次趋近律,在整个趋近阶段都具有较好的收敛速度。文献[7]提出了一种双幂次趋近律,文献[8]指出快速幂次和双幂次趋近律均具有二阶滑模特性,并推导了稳态误差界。文献[9]分析指出双幂次趋近律的二阶滑模运动在有限时间内形成,并给出了收敛时间的估计,也有研究进一步指出,双幂次趋近律具有固定时间收敛特性,并可以给出收敛时间上界。

本文在以上研究的基础上,结合快速幂次和双幂次趋近律,提出了一种新型双幂次组合函数趋近律。

1 双幂次组合趋近律设计

文献[6]和文献[7]分别提出了快速幂次趋近律和双幂次趋近律。

$$\dot{s} = -k_1 s - k_2 |s|^{1-\gamma} \text{sgn}(s) \tag{1}$$

$$\dot{s} = -k_1 |s|^{1+\gamma} \text{sgn}(s) - k_2 |s|^{1-\gamma} \text{sgn}(s) \tag{2}$$

其中, $k_1 > 0, k_2 > 0, 0 < \gamma < 1$ 。若不考虑干扰,上述两种趋近律均可以实现二阶滑模动态,即有限时间内使得 $s = \dot{s} = 0$ 。系统初始状态到达滑模面的过程分为两个阶段:当系统状态远离滑模面,即 $|s| > 1$ 时,式(1)和式(2)的等号右侧第一项起主导作用;当系统状态接近滑模面时,即 $|s| < 1$ 时,则是等号右侧第二项起主导作用。

假设初始状态满足 $s(0) = s_0 > 1$,比较上述两种趋近律的收敛时间,分两个阶段进行讨论。

1) 第一阶段: $s(0) = s_0 \rightarrow s(t_1) = 1$ 。

取 Lyapunov 函数 $V = s^2$, 结合式(1)和式(2)分别得到:

$$\dot{V} = -2k_1 V - 2k_2 V^{1-\gamma/2} \tag{3}$$

$$\dot{V} = -2k_1 V^{1+\gamma/2} - 2k_2 V^{1-\gamma/2} \tag{4}$$

对比式(3)和式(4)可以看出,只有等号右侧第一项存在差异,而在该阶段,同样是等号右侧第一项起主导作用。因此,只需分析此项就可以比较收敛时间。

忽略等号右侧第二项,分别对式(3)和式(4)两边求积分,得:

$$V(t) = V_0 \exp(-2k_1 t) \tag{5}$$

$$V^{-\gamma/2}(t) = V_0^{-\gamma/2} + \gamma k_1 t \tag{6}$$

将 $V(t_1) = s^2(t_1) = 1$ 分别代入式(5)和式(6),计算得 $s_0 \rightarrow 1$ 所需时间为:

$$t_{s1} = -\frac{\ln(1-x)}{k_1 \gamma} \tag{7}$$

$$t_{d1} = \frac{x}{k_1 \gamma} \tag{8}$$

其中: $x = 1 - V_0^{-\gamma/2}$; 根据对数函数不等式 $-\ln(1-x) > x$, 可得 $t_{s1} > t_{d1}$ 。说明在第一阶段 ($|s(t)| > 1$), 式(4)具有更快的收敛速度。

2) 第二阶段: $s(t_1) = 1 \rightarrow s(T) = 0$ 。

此时, 式(3)和式(4)的第二项起主导作用, 由于两式中第二项相同, 因此, 比较该阶段的收敛时间应同时考虑所有项。参考文献[10], 快速幂次和双幂次趋近律收敛到原点所需时间分别为:

$$T = \frac{1}{k_1 \gamma} \ln \left(1 + \frac{k_1}{k_2} |s_0|^\gamma \right) \tag{9}$$

$$T = -\frac{|s_0|^{-\gamma}}{\gamma} k_1^{-\gamma/(1+\gamma)} \cdot F \left(1, \frac{1}{2}; \frac{3}{2}; -\frac{k_2}{k_1} |s_0|^{-2\gamma} \right) \tag{10}$$

其中, $F(\cdot)$ 为高斯超几何函数, 其定义^[11]为

$$\begin{aligned} F(\alpha, \beta; \gamma; z) &= \sum_{n=0}^{\infty} \frac{(\alpha)_n (\beta)_n}{(\gamma)_n n!} z^n \\ &= 1 + \frac{\alpha \beta}{\gamma} z + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)2} z^2 + \dots \end{aligned} \tag{11}$$

其中: $\alpha, \beta, \gamma \in \mathbb{R}$, $(\alpha)_n$ 表示 α 的波赫默默 n 阶乘幂, 定义为 $(\alpha)_n = \alpha(\alpha+1)(\alpha+2)\cdots(\alpha+n-1)$, $n \in \mathbb{N}$, 特别地, $(\alpha)_0 = 1, (1)_n = n!$ 。

将该阶段初始状态 $s_0 = s(t_1) = 1$ 代入式(9)和式(10), 收敛所需时间分别为:

$$t_{s2} = \frac{1}{k_1 \gamma} \ln \left(1 + \frac{k_1}{k_2} \right) \tag{12}$$

$$t_{d2} = -\frac{1}{k_1^{\gamma/(1+\gamma)} \gamma} \cdot F \left(1, \frac{1}{2}; \frac{3}{2}; -\frac{k_2}{k_1} \right) \tag{13}$$

进一步计算 t_{d2}/t_{s2} 得到:

$$\frac{t_{d2}}{t_{s2}} = \frac{z \cdot \arctan(z)}{\ln(1+z^2)} \tag{14}$$

式(14)的推导利用了高斯超几何函数定义式(11)和 $\arctan(\cdot)$ 函数的幂级数展开式。其中 $z = \sqrt{k_1/k_2} > 0$ 。根据实函数理论, 可以推导得到 $z \cdot \arctan(z) > \ln(1+z^2)$ (如图 1(a) 所示)。因此 $t_{d2} > t_{s2}$, 说明在第二阶段 ($|s(t)| < 1$) 式(1)的收敛速度更快。

通过对上述两种趋近律的收敛时间的分析, 提出一种新型双幂次组合函数趋近律。

$$\dot{s} = -k_1 \text{fal}(s, a, \delta) - k_2 |s|^b \text{sgn}(s) \tag{15}$$

其中, $a = 1 + \gamma, b = 1 - \gamma, \delta = 1, 0 < \gamma < 1$, 非线性

性幂次组合函数 $fal(\cdot)$ (函数曲线如图 1(b) 所示) 的形式^[12] 为

$$fal(s,a,\delta)=\begin{cases} |s|^a \text{sgn}(s), & |s|>\delta \\ \frac{s}{\delta^{1-a}}, & |s|\leq\delta \end{cases} \quad (16)$$

当 $|s|>1$ 时, 式(15)等价于式(2); 而当 $|s|<1$ 时, 式(15)又等价于式(1)。根据本节前文的分析可知, 与现有的快速幂次及双幂次趋近律相比, 新型趋近律(式(15))具有更快的收敛速度。

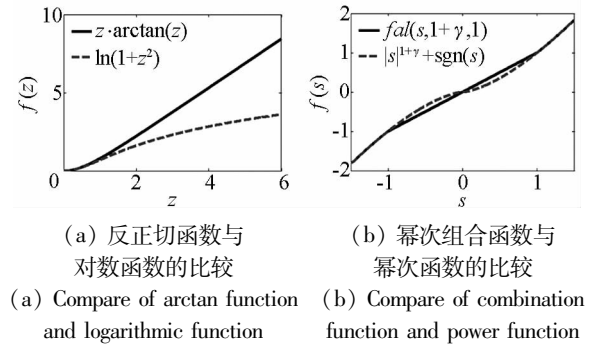


图 1 $z \cdot \arctan(z)$, $\ln(1+z^2)$, $fal(s, 1+\gamma, 1)$ 和 $|s|^{1+\gamma} \text{sgn}(s)$ 的函数曲线图

Fig. 1 Curves of functions of $z \cdot \arctan(z)$, $\ln(1+z^2)$, $fal(s, 1+\gamma, 1)$ and $|s|^{1+\gamma} \text{sgn}(s)$

2 新型趋近律特性分析

2.1 固定时间收敛特性

在给出新趋近律收敛时间特性之前, 先引入固定时间收敛的定义和引理^[13]。

定义 1 设一个系统的初始状态为 $x_0 \in \mathbb{R}^n$, 原点是全局有限时间收敛平衡点, 如果收敛时间函数 $T(x_0)$ 有界, 即存在时间常数 T_{\max} , 使得 $T(x_0) \leq T_{\max}$ 对任意初始状态 x_0 成立, 则称原点是系统的全局固定时间收敛平衡点。

引理 1 若连续的径向无界函数 $V(x): \mathbb{R}^n \rightarrow \mathbb{R}^+ \cup \{0\}$ 满足下述两个条件:

- 1) $V(0) = 0$, 原点是全局有限时间收敛平衡点;
- 2) 存在 $0 < \mu < 1$, $\nu > 0$, $r_\mu > 0$ 和 $r_\nu > 0$ 使式(17)成立。

$$\dot{V} \leq \begin{cases} -r_\mu V^{1-\mu}, & V \leq 1 \\ -r_\nu V^{1+\nu}, & V > 1 \end{cases} \quad (17)$$

则原点是全局固定时间收敛平衡点, 最大收敛时间为:

$$T_{\max} = \frac{1}{\mu r_\mu} + \frac{1}{\nu r_\nu} \quad (18)$$

根据引理 1, 可以证明新型双幂次组合函数

趋近律(式(15))满足定理 1。

定理 1 对式(15), 状态 (s, \dot{s}) 在固定时间 T_{\max} 内收敛到 0, 即在有限时间 $T(s_0)$ 后有 $s = \dot{s} = 0$, 收敛时间 $T(s_0)$ 存在与初始状态 s_0 无关的上界。

$$T_{\max} = \frac{1}{\gamma} \left(\frac{1}{k_2} + \frac{1}{k_1} \right) \quad (19)$$

证明: 选取 Lyapunov 函数

$$V = s^2 \quad (20)$$

显然, 式(20)满足 $V(0) = 0$ 。

当状态 $s(t)$ 在区域 $|s|>1$ 中时, 式(15)等价于式(2)。根据文献[9]可知, 式(2)的原点是全局有限时间收敛平衡点。可以推论: 在有限时间内状态 $s(t)$ 收敛到区域 $|s|<1$ 中, 此时式(15)等价于式(1), 状态 $s(t)$ 收敛到 0 的时间满足式(7), 可知收敛时间是有限的。因此对于式(15), $s = 0$ 是全局有限时间收敛的平衡点。引理 1 的第一个条件成立。

对 V 沿式(15)轨迹求导, 得:

$$\begin{aligned} \dot{V} &= 2s\dot{s} \\ &= \begin{cases} -2k_1 V - 2k_2 V^{1-\gamma/2}, & V \leq 1 \\ -2k_1 V^{1+\gamma/2} - 2k_2 V^{1-\gamma/2}, & V > 1 \end{cases} \\ &\leq \begin{cases} -2k_2 V^{1-\gamma/2}, & V \leq 1 \\ -2k_1 V^{1+\gamma/2}, & V > 1 \end{cases} \end{aligned} \quad (21)$$

满足引理 1 的第二个条件。对比式(21)与式(17), 参数对应关系为: $r_\mu = 2k_2$, $\mu = \nu = 0.5\gamma$, $\nu = 2k_1$ 。综上所述, 结合引理 1 可知, $s = 0$ 是式(15)的全局固定时间收敛平衡点。收敛时间 $T(s_0)$ 满足:

$$T(s_0) \leq T_{\max} = \frac{1}{\mu r_\mu} + \frac{1}{\nu r_\nu} = \frac{1}{\gamma} \left(\frac{1}{k_2} + \frac{1}{k_1} \right)$$

至此定理得证。□

2.2 稳态误差界分析

考虑式(15)受到不确定扰动 d 的影响, 系统方程变为:

$$\dot{s} = -k_1 fal(s, 1+\gamma, 1) - k_2 |s|^{1-\gamma} \text{sgn}(s) + d \quad (22)$$

式中, 不确定扰动 d 未知但有界, 即 $|d| \leq D$ 。式(22)的稳态误差界满足定理 2。

定理 2 式(22)的状态 s 在有限时间内收敛到以下区域。

$$|s| \leq \min \{ D/k_1, (D/k_2)^{\frac{1}{1-\gamma}}, (D/k_1)^{\frac{1}{1+\gamma}} \}$$

证明: 选择 Lyapunov 函数

$$V = 0.5s^2 \quad (23)$$

沿式(22)轨线求得:

$$\dot{V} = -k_1 fal(s, 1 + \gamma, 1) s - k_2 |s|^{2-\gamma} + d \cdot s$$

上式可进一步写成以下四种形式:

$$\dot{V} \leq \begin{cases} -k_2 |s|^{2-\gamma} - |s|(k_1 |s| - D) \\ -k_1 |s|^2 - |s|(k_2 |s|^{1-\gamma} - D) \end{cases}, |s| \leq 1 \quad (24)$$

$$\dot{V} \leq \begin{cases} -k_2 |s|^{2-\gamma} - |s|(k_1 |s|^{1+\gamma} - D) \\ -k_1 |s|^{2+\gamma} - |s|(k_2 |s|^{1-\gamma} - D) \end{cases}, |s| > 1 \quad (25)$$

1) 当 $1 \geq |s| \geq D/k_1$, 即 $0.5 \geq V \geq V_1 = 0.5(D/k_1)^2$ 时, 由式(24)第一式可知:

$$\dot{V} \leq -k_2 |s|^{2-\gamma} = -2^{1-\gamma/2} k_2 V^{1-\gamma/2} \quad (26)$$

注意到 $1 - \gamma/2$ 小于 1。说明如果 $D/k_1 < 1$, 则有限时间内系统收敛到区域 $|s| \leq D/k_1$ 中。

2) 当 $1 \geq |s| \geq (D/k_2)^{\frac{1}{1-\gamma}}$, 即 $0.5 \geq V \geq V_1 = 0.5(D/k_2)^{\frac{2}{1-\gamma}}$ 时, 由式(24)第二式可知:

$$\dot{V} \leq -k_1 |s|^2 = -2k_1 V \quad (27)$$

从 $V_0 = 0.5s_0^2$ 收敛到 V_1 所需时间 $T \leq T_{\max}$, 其中 $T_{\max} = 0.5k_1^{-1} \ln(V_0/V_1)$, 说明如果 $D/k_2 < 1$, 则有限时间内系统收敛到区域 $|s| \leq (D/k_2)^{\frac{1}{1-\gamma}}$ 中。

3) 当 $|s| > (D/k_1)^{\frac{1}{1+\gamma}} > 1$, 即 $0.5 \geq V \geq V_1 = 0.5(D/k_1)^{\frac{2}{1+\gamma}}$ 时, 由式(25)第一式可知:

$$\dot{V} \leq -k_2 |s|^{2-\gamma} = -2^{1-\gamma/2} k_2 V^{1-\gamma/2}$$

与式(26)相同, 说明如果 $D/k_1 > 1$, 则有限时间内系统收敛到区域 $|s| \leq (D/k_1)^{\frac{1}{1+\gamma}}$ 中。

4) 当 $|s| > (D/k_2)^{\frac{1}{1-\gamma}} > 1$, 即 $0.5 \geq V \geq V_1 = 0.5(D/k_2)^{\frac{2}{1-\gamma}}$ 时, 由式(25)第二式可知:

$$\dot{V} \leq -k_1 |s|^{2+\gamma} = -2^{1+\gamma/2} k_1 V^{1+\gamma/2} \quad (28)$$

从 $V_0 = 0.5s_0^2$ 收敛到 V_1 所需时间 $T \leq T_{\max}$, 其中 $T_{\max} = \frac{1}{2^{\gamma/2} k_1 \gamma} \left(\frac{1}{V_1^{\gamma/2}} - \frac{1}{V_0^{\gamma/2}} \right)$, 说明如果 $D/k_2 > 1$, 则有限时间内系统收敛到区域 $|s| < (D/k_2)^{\frac{1}{1-\gamma}}$ 中。

综上所述, 状态 s 将在有限时间内收敛到如式(29)所示区域。

$$|s| \leq \min\{D/k_1, (D/k_1)^{\frac{1}{1+\gamma}}, (D/k_2)^{\frac{1}{1-\gamma}}\} \quad (29)$$

至此定理得证。□

注释 1 文献[8]给出了式(1)和式(2)的稳态误差界, 分别为:

$$|s| \leq \min\{D/k_1, (D/k_2)^{\frac{1}{1-\gamma}}\} \quad (30)$$

$$|s| \leq \min\{(D/k_1)^{\frac{1}{1+\gamma}}, (D/k_2)^{\frac{1}{1-\gamma}}\} \quad (31)$$

根据式(29)~(31)可见, 如果 $(D/k_2)^{\frac{1}{1-\gamma}}$ 最小, 三种趋近律的稳态误差界相同; 如果 $(D/k_2)^{\frac{1}{1-\gamma}}$ 最大, 则稳态误差界取决于 D/k_1 , 当 $D/k_1 > 1$ 时, $(D/k_1)^{\frac{1}{1+\gamma}} < D/k_1$, 而当 $D/k_1 < 1$ 时, $(D/k_1)^{\frac{1}{1+\gamma}} > D/k_1$ 。提出的幂次组合函数趋近律在任何情况下($D/k_1 \in (0, +\infty)$)的稳态误差总是小于或等于现有的快速幂次或双幂次趋近律。

3 仿真算例

考虑单输入单输出系统

$$\dot{s} = u + d(t) \quad (32)$$

式中, u 为控制输入, $d(t)$ 为时变不确定扰动。分别利用快速幂次趋近律、双幂次趋近律和提出的双幂次组合函数趋近律设计控制律 u 并进行仿真。控制参数取为 $k_1 = 4, k_2 = 1, \gamma = 0.5$ 。利用各趋近律设计的控制律为:

1) 快速幂次趋近律

$$u_1 = -4s - |s|^{0.5} \text{sgn}(s) \quad (33)$$

2) 双幂次趋近律

$$u_2 = -4|s|^{1.5} \text{sgn}(s) - |s|^{0.5} \text{sgn}(s) \quad (34)$$

3) 双幂次组合函数趋近律

$$u_3 = -4 \cdot fal(s, 1.5, 1) - |s|^{0.5} \text{sgn}(s) \quad (35)$$

3.1 收敛时间仿真对比

当式(32)不存在扰动, 即 $d(t) = 0$ 时, 设置初始状态分别为 $s_0 = 1, s_0 = 10, s_0 = 100$, 控制输入 u 分别采用式(33)~(35)所示控制律, 对比三种趋近律的收敛时间。状态变量 s 的时间历程曲线如图 2 所示, 图 2 中纵轴采用对数坐标。

图 2(a)表明, 当初始状态 $s_0 = 1$ 时, 所提趋近律与快速幂次趋近律收敛速度相同, 收敛时间约为 0.8 s, 均小于双幂次趋近律收敛时间(约 1.1 s)。图 2(b)中, 当初始状态 $s_0 = 10$ 时, 所提趋近律具有最快收敛速度, 收敛时间约为 1.1 s, 快速幂次趋近律次之, 收敛时间约为 1.3 s, 双幂次趋近律收敛最慢, 收敛时间为 1.4 s。图 2(c)显示, 初始状态 $s_0 = 100$ 时, 所提趋近律收敛速度仍为最快, 收敛时间约为 1.2 s, 而快速幂次趋近律收敛速度最慢, 收敛时间约为 1.85 s, 双幂次趋近律收敛时间为 1.5 s。

对比图 2 中的 3 个子图, 可以看出快速幂次趋近律收敛时间受初始状态影响很大, 不能看出存在收敛时间上限, 而双幂次趋近律和所提双幂次组合函数趋近律的收敛时间受初始状态变化影

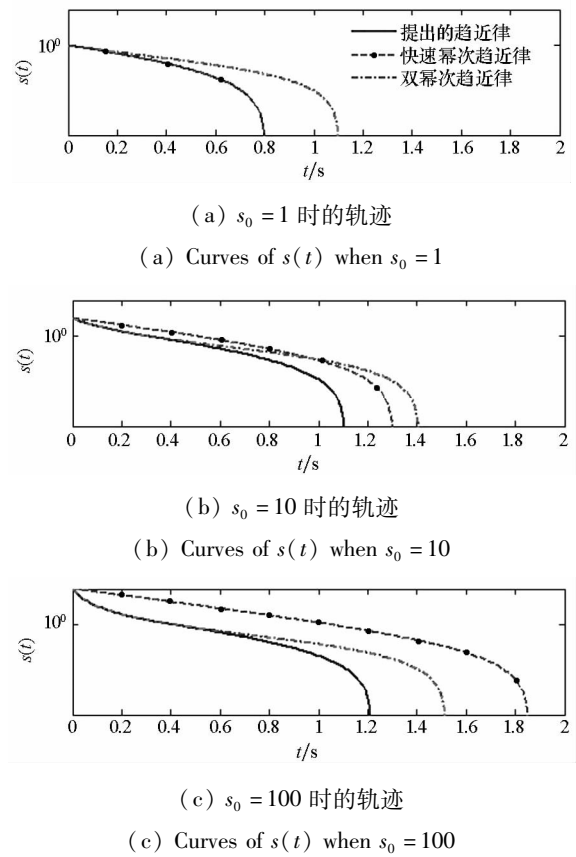


图2 不同控制律 u 作用下 s 的收敛曲线

Fig.2 Convergence curves of s by different control inputs u

响较小。根据定理1,存在与初始状态无关的收敛时间上限。在本例中,理论上的最大收敛时间 $T_{\max}=2.5$ s。与现有的双幂次趋近律相比,所提趋近律不仅保持了收敛时间上限的存在,还使实际收敛速度更快。

3.2 稳态误差界仿真对比

当式(32)存在扰动,即 $d(t) \neq 0$ 时,设置初始状态 $s_0=6$,分别采用式(33)~(35)所列控制律进行仿真,以对比不同趋近律的稳态误差界。取扰动项 $d(t)$ 为以下两种不同的情况。

1) 扰动上界 $D=10$ 时,

$$d_1(t) = 7\sin(2t) + 3\cos t$$

根据式(29)~(31),计算所提趋近律和双幂次趋近律的稳态误差上界为 1.842 0,快速幂次趋近律的稳态误差上界为 2.500 0。

2) 扰动上界 $D=1$ 时,

$$d_2(t) = 0.3\cos(2t) + 0.7\sin t$$

计算得出所提趋近律和快速幂次趋近律的稳态误差上界为 0.250 0,双幂次趋近律的稳态误差上界为 0.396 9。

仿真结果如图3和图4所示。可见,在受扰情况下,状态 s 没有收敛到0,而是有限时间收敛

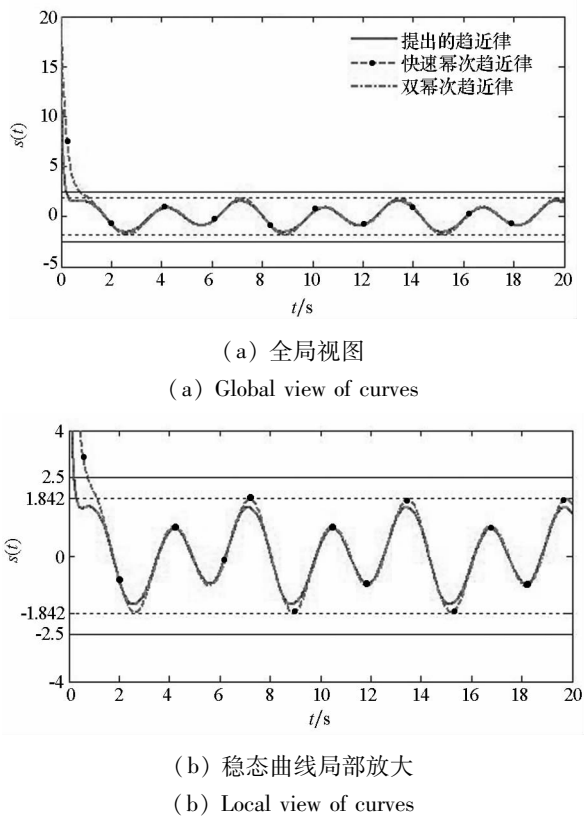


图3 扰动上界 $D=10$ 时的稳态误差曲线

Fig.3 Curves of stabilized error when disturbance upper bound $D=10$

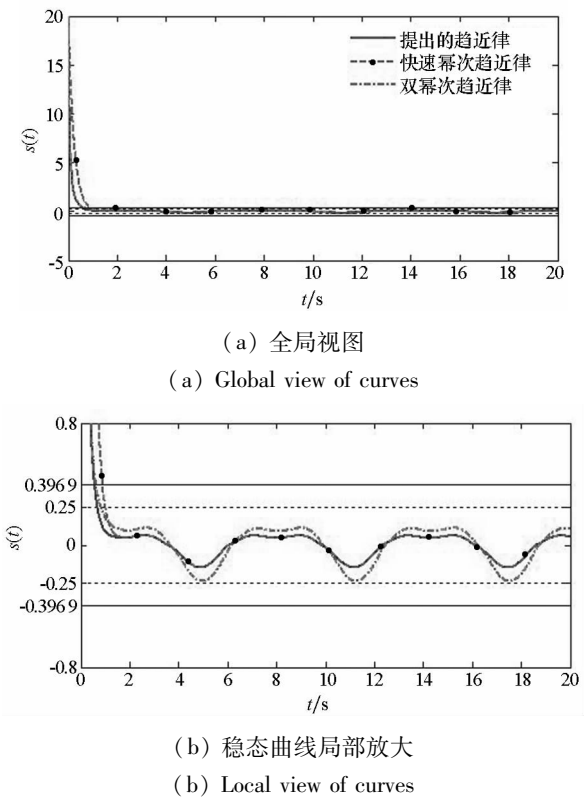


图4 扰动上界 $D=1$ 时的稳态误差曲线

Fig.4 Curves of stabilized error when disturbance upper bound $D=1$

到稳态误差界之内,此后误差轨迹不再超出式(29)~(31)所描述的范围。图 3 显示,当不确定扰动为 $d_1(t)$ 时,所提趋近律与双幂次趋近律的稳态轨迹相同,实际误差范围小于快速幂次趋近律。图 4 表明,不确定扰动为 $d_2(t)$ 时,所提趋近律又与快速幂次趋近律的稳态轨迹相同,实际误差范围小于双幂次趋近律。综合图 3 与图 4 的仿真结果可知,式(29)是正确的。同时,与现有的两种趋近律相比,所提新型趋近律对任意有界扰动具有更好的稳态品质。

4 结论

本文提出了一种双幂次组合函数趋近律设计方案,相比现有趋近律,具有收敛速度快、稳态误差小的特点,还能够解决滑模控制中的抖振问题。理论分析表明:所提新型趋近律无论是在远离还是接近滑动模态时都具有很快的趋近速度,不仅收敛总时间小于现有的快速幂次及双幂次趋近律,能在有限时间内实现二阶滑模,即 $s = \dot{s} = 0$ 。还具有固定时间收敛的特性,即有限收敛时间的上界与初始状态无关;当存在有界外扰时,状态 s 收敛于平衡零点的有界领域内,同时稳态误差范围也小于现有快速幂次趋近律或双幂次趋近律。将该趋近律与滑模干扰观测器结合,可以实现无抖振快速连续控制。当双幂次组合函数趋近律取更一般的形式($a > 1; 0 < b < 1; k_1, k_2, \delta > 0$)时,滑模收敛特性有待进一步的研究。

参考文献 (References)

- [1] Lee H, Utkin V I. Chattering suppression methods in sliding mode control systems[J]. Annual Reviews in Control, 2007, 31(2): 179–188.
- [2] Fridman L, Levant A. Higher order sliding modes[J]. Sliding Mode Control in Engineering, 2002(11): 53–102.
- [3] Wang H, Han Z Z, Xie Q Y. Finite-time chaos control via nonsingular terminal sliding mode control [J]. Communications in Nonlinear Science and Numerical Simulation, 2009, 14(6): 2728–2733.
- [4] Koshkouei A J, Burnham K J. Dynamic sliding mode control design [J]. IEEE Proceedings Control Theory and Applications, 2005, 152(4): 392–396.
- [5] 高为炳. 变结构控制的理论及设计方法[M]. 北京: 科学出版社, 1996.
- GAO Weibing. Theory and design method for variable sliding mode control [M]. Beijing: Science Press, 1996. (in Chinese)
- [6] Yu S H, Yu X H, Shirinzadeh B, et al. Continuous finite-time control for robotic manipulators with terminal sliding mode[J]. Automatica, 2005, 41(11): 1957–1964.
- [7] 梅红, 王勇. 快速收敛的机器人滑模变结构控制[J]. 信息与控制, 2009, 38(5): 552–557.
- MEI Hong, WANG Yong. Fast convergent sliding mode variable structure control of robot [J]. Information and Control, 2009, 38(5): 552–557. (in Chinese)
- [8] 李鹏, 马建军, 郑志强. 采用幂次趋近律的滑模控制稳态误差界[J]. 控制理论与应用, 2011, 28(5): 619–624.
- LI Peng, MA Jianjun, ZHENG Zhiqiang. Sliding mode control approach based on nonlinear integrator[J]. Control Theory and Applications, 2011, 28(5): 619–624. (in Chinese)
- [9] 张合新, 范金锁, 孟飞, 等. 一种新型滑模控制双幂次趋近律[J]. 控制与决策, 2013, 28(2): 289–293.
- ZHANG Hexin, FAN Jinsuo, MENG Fei, et al. A new double power reaching law for sliding mode control [J]. Control and Decision, 2013, 28(2): 289–293. (in Chinese)
- [10] Yang L, Yang J Y. Nonsingular fast terminal sliding-mode control for nonlinear dynamical systems [J]. International Journal of Robust and Nonlinear Control, 2011, 21(16): 1865–1879.
- [11] Olver F W, Lozier D W, Boisvert R F, et al. NIST handbook of mathematical functions [M]. US: Cambridge University Press, 2010.
- [12] Han J. From PID to active disturbance rejection control[J]. IEEE Transactions on Industrial Electronics, 2009, 56(3): 900–906.
- [13] Polyakov A, Fridman L. Stability notions and Lyapunov functions for sliding mode control systems[J]. Journal of the Franklin Institute, 2014, 351(4): 1831–1865.

多级可修备件库存的生灭过程建模与优化*

刘任洋¹, 李 华¹, 李庆民², 熊宏锦³

(1. 海军工程大学 兵器工程系, 湖北 武汉 430033; 2. 海军工程大学 科研部, 湖北 武汉 430033;
3. 海军装备部驻重庆地区军事代表局, 重庆 400042)

摘 要:针对 VARI-METRIC 模型在低可用度下结果不准确的问题, 建立基于生灭过程的任意等级、任意层级可修件库存优化模型。通过对各级站点、各类备件需求率与到达率的预测, 对每个部件建立其生灭过程模型, 并提出基于生灭过程的装备可用度计算方法。以整个保障系统的装备可用度为约束指标, 以备件总购置费最低为目标, 利用边际算法得到最优备件配置方案, 并建立仿真模型对所得优化方案进行评估与调整。结合算例, 以仿真结果作为检验标准, 选取权威的 VMETRIC 软件与该解析模型在优化性能、计算精度及适用性上进行对比和说明。结果表明, 无论是解析模型还是 VMETRIC 软件, 均存在一定的适用范围, 而采用解析与仿真相结合的方法无疑具有更强的适应性。

关键词:生灭过程; 可修复备件; 可用度; 库存优化
中图分类号:E911; TJ761.1; V125.7 **文献标志码:**A **文章编号:**1001-2486(2017)03-111-10

Modeling and optimization of multi-echelon inventory for repairable spares based on birth and death process

LIU Renyang¹, LI Hua¹, LI Qingmin², XIONG Hongjin³

(1. Department of Weaponry Engineering, Naval University of Engineering, Wuhan 430033, China;
2. Office of Research & Development, Naval University of Engineering, Wuhan 430033, China;
3. Military Representative Office of Naval Equipment Department in Chongqing, Chongqing 400042, China)

Abstract: For it is not accurate under the condition of low availability, VARI-METRIC model of inventory optimization for multi-echelon multi-indenture repairable spares was built. Firstly, the birth and death process of each component was established by the prediction of demand rate and arrival rate of each spares in each site. Then, a computational method of availability was put forward based on the birth and death process. With the constraints of availability and objective of lowest cost, the optimal inventory distribution result was obtained by marginal algorithm and the simulation model was built to evaluate and adjust the result. In an actual example, the analytic model and the VMETRIC were compared and described in aspects of optimization performance, calculation precision and applicability by simulation verification. Results show that both the analytic model and the VMETRIC have certain scope of applicability and the method combined analytic model and simulation has a stronger applicability.

Key words: birth and death process; repairable spares; availability; inventory optimization

可修复性备件的配置问题是备件规划工作的重要环节。多级维修供应是较为科学的保障模式, 目前国内外各军兵种大都采用该模式。由于装备使用现场的维修条件和备件储备能力有限, 因此维修、备件储备及供应等保障活动在各级站点之间协调进行。从装备的全寿命周期角度看, 由于可以得到包括工业部门或外部供应商在内的所有保障组织体系的支持, 备件在供应过程中一般不存在实质性的消耗, 具体表现为顶层站点具

备较强的维修能力, 能对所有故障件进行完全修复, 或即使因无法修复而报废但能通过采购方式得到补充。在这种没有实质消耗的情况下, 备件在长期的维修、补给过程中, 其供应渠道数量、库存概率将趋于稳定^[1]。由此可以看出, 稳态条件下各级保障站点的库存配置问题是对装备及其备件从列装到退役整个寿命周期的总体规划, 对于军方掌控和把握新列装备所需的配套备件具有重要意义。

* 收稿日期:2016-01-12
基金项目:国家部委基金资助项目(51304010206, 51327020105)
作者简介:刘任洋(1989—), 男, 江西南昌人, 博士研究生, E-mail:463572090@qq.com;
李庆民(通信作者), 男, 教授, 博士, 博士生导师, E-mail:licheng001@hotmail.com

对于稳态条件下可修件多级库存问题的研究,目前主流的解析建模方法是采用可修复备件多级库存控制技术(Multi-Echelon Technology for Recoverable Item Control, METRIC)系列模型理论,包括 METRIC^[2]、MOD-METRIC^[3]、VARI-METRIC^[4]等模型。其中,METRIC 模型是该系列模型的基础,而 VARI-METRIC 模型是最终形式的多级保障结构、多层级备件配置优化模型。国内外诸多学者基于 VARI-METRIC 模型,并结合实际保障需求对模型进行扩展和改进,解决了一系列诸如有限维修渠道^[5-7]、串件拼修^[8-10]、横向补给^[11-14]等情况下的备件方案评估和优化问题。国外较为先进的备件优化工具 VMETRIC, OPUS10 均将其作为核心模型与算法。然而由于 VARI-METRIC 模型在建模过程中存在一些近似与假设,其结果的准确性和精度如何,没有公开的文献对其进行全面系统的验证。除了 VARI-METRIC 建模思想,是否还有其他效果更好或互补的适用于多级库存的建模方法也是值得研究和探讨的问题。

1 保障过程描述及模型说明

1.1 多级保障过程描述

装备一般包含多个结构层次,根据在装备系统所处的不同结构层次,备件分为现场可更换单元(Line Replacable Unit, LRU)、车间更换单元(Shop Replacable Unit, SRU)等,图 1 所示为一个典型的多层次结构系统。

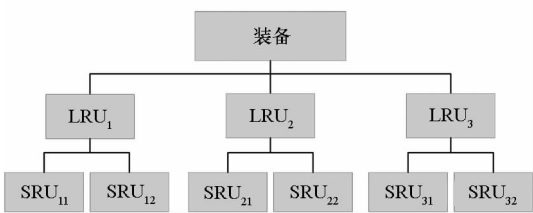


图 1 装备层次结构

Fig. 1 Hierarchical structure of equipment

装备在工作时发生故障而停机,是所属的第一层级部件 LRU 故障导致的,采用换件维修的方式将故障 LRU 拆卸。如果现场有该 LRU 备件,则立刻进行更换完成装备的修理,如果没有 LRU 备件,就发生一次 LRU 备件短缺。受维修条件限制,拆下的故障 LRU 以一定的概率在现场站点修复成功,如果现场站点不能维修,则送向上级保障站点维修并向上级申领一项该备件。在对故障 LRU 进行维修时,故障的原因是其所属的 SRU 故

障导致的;如果有该 SRU 备件,则将其安装到 LRU 上,从而完成对 LRU 的修理;如果没有 SRU 备件,则需等待 SRU 的维修,从而造成 LRU 的修理延误。故障 SRU 在各级站点也存在一定的修复概率,其送修和申请补给过程与 LRU 相同。当完成了一件 LRU 的修理或补给时,备件短缺事件就得以解决。图 2 为一个典型的三级保障体系,由一个基地级站点 b_0 ,两个中继级站点 n_1 、 n_2 和三个现场站点 j_1 、 j_2 、 j_3 构成。

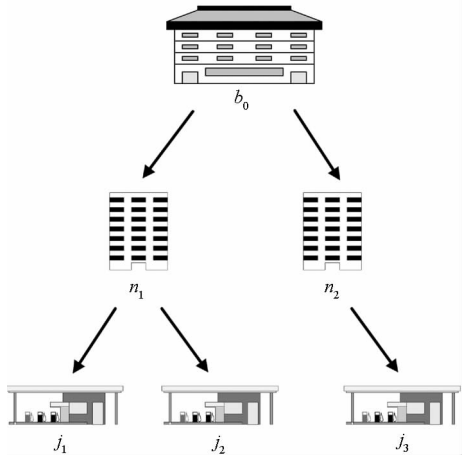


图 2 三级保障组织结构

Fig. 2 Support structure of three-echelon

1.2 模型假设及参数定义

为了简化建模过程,在上述保障过程描述的基础上做出如下几点假设和说明:

- 1) 所有备件需求率均服从泊松分布;
- 2) 各级站点均采用连续检测的 $(S-1, S)$ 库存策略,即缺少一件就向上级申请一件;
- 3) 各级站点维修时间服从指数分布,维修渠道不限,不考虑故障单元之间的维修优先权,且忽略重测完好率、虚警率等维修参数的影响;
- 4) LRU 的故障只是由于其所属 SRU 的某一故障所致,不考虑多个 SRU 同时故障的情况;
- 5) 顶层站点具有较强的修复能力,能对所有故障件进行完全维修,不考虑报废问题;
- 6) 部署于现场站点的多台同型装备之间为独立关系,工作时相互不受影响;装备中同一层级部件之间为串联关系。

模型相关参数定义如下:

j :保障站点编号($j=1,2,\cdots,J$), J 表示保障站点总数;

h :站点的级别编号($h=1,2,\cdots,H$), $h=1$ 表示顶层站点, $h=H$ 表示底层站点(舰员级), $h=2,3,\cdots,H-1$ 表示处于中间级别站点;

$Echelon(h)$:处于第 h 个级别的站点集合;

$Unit(j)$:站点 j 保障的所有下一级别站点集合;

$Sup(j)$:站点 j 的上级站点;

i :部件项目编号 ($i = 1, 2, \dots, I$), I 表示部件类型总数;

c :部件层级编号 ($c = 0, 1, \dots, C$), $c = 0$ 表示装备系统, $c = 1$ 表示第一层级部件 LRU, $c = C$ 表示处于装备中最底层部件, $c = 2, 3, \dots, C - 1$ 表示处于中间结构层级部件;

$Inden(c)$:在装备结构中处于第 c 层次的项目集合;

$Sub(i)$:部件 i 所属下一层级的分组件集合;

$Aub(i)$:部件 i 上面所有层级的母体集合;

$MTBF_i$:部件 i 的平均故障间隔时间;

q_{iz} :部件 i 发生故障是由于其所属于部件 z 故障导致的条件概率 ($z \in Sub(i)$);

S_{ij} :站点 j 第 i 项备件的库存量;

Td_{ij} :站点 j 第 i 项备件的平均短缺时间;

EBO_{ij} :站点 j 第 i 项备件的期望短缺数;

λ_{ij} :单装备下站点 j 第 i 项备件的需求率;

μ_{ij} :单维修渠道下站点 j 第 i 项备件的到达率;

$P_{S_{ij}}^{ij}$:站点 j 第 i 项备件库存量为 S 时的稳态概率;

T_{ij} :部件 i 在站点 j 的平均维修时间;

r_{ij} :部件 i 在站点 j 的维修概率;

O_{ij} :站点 j 第 i 项备件的补给运输时间;

Z_i :部件 i 在其母体中的单机安装数量;

τ_i :备件 i 的单价;

M_j :装备的站点 j 的部署数量;

HD_j :装备在站点 j 平均每天工作时间;

A_j :站点 j 的装备可用度;

2 基于近似生灭过程的多级稳态库存建模

2.1 可修件的生灭过程模型

生灭过程是更新过程的一种特例,其特征是:在很短的时间内,处于状态 s 的系统只能转移到状态 $s + 1$ 或 $s - 1$ 或保持不变^[15]。其中,系统从状态 s 转移到状态 $s - 1$ 的概率称为死亡率;系统从状态 s 转移到状态 $s + 1$ 称为出生率。如果将备件的现有库存数量 S 定义为系统的状态,假设维修渠道不限,就能建立可修件的生灭过程模型。由于考虑各装备之间为独立关系,现场站点的备件库存数量最低为 $-M$ (短缺 M 件),其生灭过程模型如图3所示。 $-M, \dots, -1$ 均为缺件状态,也表示装备停机数量。 λ 为单装备下的备件需求

率, μ 为单维修渠道下的备件到达率。对于单站点,备件到达即指故障件修复成功;对于多级站点,备件到达除了指本站点故障件修复成功外,还应包括上级站点备件的补给成功。

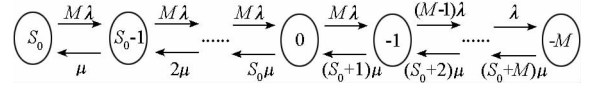


图3 可修件的生灭过程模型

Fig. 3 Birth and death process model for repairable spares

定义 P_s 表示库存数量为 S 的稳态概率,可列出生灭过程的状态转移方程为:

$$M\lambda P_{S_0} = \mu P_{S_0-1} \quad (1)$$

$$(M\lambda + k\mu) P_{S_0-k} = M\lambda P_{S_0-k+1} + (k+1)\mu P_{S_0-k-1}, \quad k = 1, 2, \dots, S_0 \quad (2)$$

$$[(M-k+S_0)\lambda + k\mu] P_{S_0-k} = (M-k+S_0+1)\lambda P_{S_0-k+1} + (k+1)\mu P_{S_0-k-1}, \quad k = S_0+1, \dots, S_0+M \quad (3)$$

$$\lambda P_{-M+1} = (S_0+M)\mu P_{-M} \quad (4)$$

结合正则条件 $\sum_{S=-M}^{S_0} P_S = 1$ 求解上述方程,可以得到系统在任意库存量下的稳态概率为:

$$P_{S_0} = \left(1 + \sum_{k=1}^{S_0+1} \frac{(M\lambda)^k}{\mu^k k!} + \sum_{k=S_0+2}^{S_0+M} \frac{M^{(S_0+1)} (M-1)! \lambda^k}{\mu^k k! (M-S_0-k)!} \right)^{-1} \quad (5)$$

$$P_{S_0-k} = \begin{cases} \frac{(M\lambda)^k}{\mu^k k!} P_{S_0} & k = 1, 2, \dots, S_0+1 \\ \frac{M^{(S_0+1)} (M-1)! \lambda^k}{\mu^k k! (M-S_0-k)!} P_{S_0} & k = S_0+2, \dots, S_0+M \end{cases} \quad (6)$$

则期望备件短缺数为:

$$EBO = \sum_{k=1}^M k P_{-k-M} \quad (7)$$

以上建模过程均针对现场站点(底层站点)的备件,对于上级站点,其生灭过程的缺件状态数为该站点所属下级站点的缺件数之和。通过以上分析,对于任意级别站点的任意层级部件,如果可以求得单装备备件需求率 λ 和单维修渠道下的备件到达率 μ ,则可利用生灭过程建立稳态条件下的多级库存模型。然而需要指出的是,生灭过程的应用条件为出生率和死亡率均服从泊松分布,由于假设部件寿命和维修时间均服从指数分布,因此对于单部件而言,其严格满足生灭过程的应用条件,而对于多等级、多层级部件,由于需考虑运输时间、维修延误等因素的影响,其出生率(备件到达率)不再服从泊松分布。故本文利用

生灭过程建模为一种近似方法。

2.2 备件需求率预测

备件需求率的计算需要综合考虑保障站点的级别和备件所处的结构层次。不考虑重测完好率、虚警率等维修参数的影响,对于底层站点 j 的顶层部件 $LRU_i (j \in Echelon(H), i \in Inden(1))$, 其备件需求率为:

$$\lambda_{ij} = \frac{HD_j Z_i}{24 \cdot MTBF_i} \quad (8)$$

对于其他层级部件 $i (i \notin Inden(1))$, 须由维修其母体组件 $l (l \in Aub(i))$ 产生。在所有需要维修的母体组件 l 中, 只有一部分比例 q_{li} 会产生对 i 的需求, 则底层站点 j 对备件 $i (j \in Echelon(H), i \notin Inden(1))$ 的需求率为:

$$\lambda_{ij} = \lambda_{lj} r_{lj} q_{li} \quad (9)$$

其中, 由维修母体组件 l 产生对 i 的需求的概率 q_{li} 可根据平均故障间隔时间计算得到, 即:

$$q_{li} = MTBF_l / MTBF_i \quad (10)$$

对于其他等级站点 $j (j \notin Echelon(H))$, 顶层部件 $LRU_i (i \in Inden(1))$ 的需求由 j 所属的下一级别站点 $d (d \in Unit(j))$ 因无法维修故障件 i 而需要向 j 送修和申领备件产生, 则有:

$$\lambda_{ij} = \sum_{d \in Unit(j)} \lambda_{id} \cdot r_{id} \quad (11)$$

对于站点 j 的其他层级备件 $i (j \notin Echelon(H), i \notin Inden(1))$, 需考虑除了下级站点对备件 i 的申领外, 还需考虑对其母体 l 维修时产生的对 i 的需求, 即:

$$\lambda_{ij} = \sum_{d \in Unit(j)} \lambda_{id} \cdot r_{id} + \sum_{l \in Aub(i)} \lambda_{lj} r_{lj} q_{li} \quad (12)$$

由此, 按照式(8)~(12)逐项递推计算可以得到整个保障体系中各级站点各项备件的需求率, 其中站点递推从底层开始, 备件递推从顶层开始。

2.3 备件到达率预测

对于单站点单层级部件而言, 备件到达率仅与故障件的维修时间有关; 当部件扩展为多层级时, 备件到达率除了考虑故障件的维修时间外, 还需考虑由于缺少子备件而产生的维修延误时间; 当保障站点扩展为多等级时, 备件到达率则还需考虑上级站点的补给时间和补给延误时间。因此, 与备件需求率类似, 备件到达率的计算也需考虑保障站点的级别以及备件所处的结构层次。但与备件需求率不同的是, 备件到达率的递推计算过程由顶层站点的底层部件开始。

对于顶层站点 j 的底层部件 $i (j \in Echelon(1), i \in Inden(C))$, 由于始终能修复成功, 其备件到达

率为:

$$\mu_{ij} = \frac{1}{T_{ij}} \quad (13)$$

对于其他层级部件 $i (j \in Echelon(1), i \notin Inden(C))$, 除了自身的维修外, 还需考虑由于缺少子备件而产生的维修延误, 备件到达率为:

$$\mu_{ij} = \frac{1}{T_{ij} + \sum_{z \in Sub(i)} q_{iz} \cdot Td_{zj}} \quad (14)$$

$$Td_{zj} = \frac{EBO_{zj}}{\lambda_{zj}(1 - EBO_{zj}/M_j)} \quad (15)$$

式(15)的分母项因子 $(1 - EBO_{zj}/M_j)$ 为需求率 λ_{zj} 的修正因子。因为由 2.2 节计算得到的备件需求率是在装备持续正常工作下的需求率, 而实际上当备件发生短缺而导致装备停机时, 在停机期间备件需求不会发生。故以 $\lambda_{zj}(1 - EBO_{zj}/M_j)$ 代替 λ_{zj} 作为对需求率的修正, 后文均按此法处理。

对于非顶层站点 $j (j \notin Echelon(1))$, 部件以 r_{ij} 的概率在本站点修复成功, 以 $(1 - r_{ij})$ 的概率由上级站点 $u (u \in Sup(j))$ 完成补给, 因此, 对于底层备件 $i (i \in Inden(C))$, 其到达率为:

$$\mu_{ij} = \frac{1}{r_{ij} T_{ij} + (1 - r_{ij})(O_{ij} + Td_{iu})} \quad (16)$$

$$Td_{iu} = \frac{EBO_{iu}}{\lambda_{iu}(1 - EBO_{iu}/\sum_{j \in Unit(u)} M_j)} \quad (17)$$

其中, Td_{iu} 为备件 i 在上级站点 u 的平均短缺时间, 视为对站点 j 的补给延误。

对于其他层级备件 $i (j \notin Echelon(1), i \notin Inden(C))$, 其到达率为:

$$\mu_{ij} = \frac{1}{r_{ij} T_{ij} + (1 - r_{ij})(O_{ij} + Td_{iu} + \sum_{z \in Sub(i)} q_{iz} \cdot Td_{zj})} \quad (18)$$

2.4 装备可用度的计算

通过备件需求率、到达率的预测及生灭过程建模, 最终得到底层站点 $j (j \in Echelon(H))$ 顶层 $LRU_i (i \in Inden(1))$ 的期望短缺数为:

$$EBO_{ij} = \sum_{k=1}^{M_j} k P_{-k}^{ij} \quad (19)$$

VARI-METRIC 模型在求出 LRU 的期望短缺数(可用度)后, 直接利用各 LRU 可用度的乘积计算装备可用度, 这将导致结果偏小。本节对装备再次利用生灭过程建模, 如图 4 所示, 由于装备不配备件, 因此对装备的建模包含 $M_j + 1$ 个状态: 0 代表所有装备正常工作, -1 代表有一台装备停机, $-M_j$ 代表所有 M_j 台装备均停机。其需求率 μ_{0j} 和到达率 λ_{0j} 分别为:

$$\mu_{0j} = \sum_{i \in \text{Inden}(1)} q_{0i} \frac{EBO_{ij}}{\lambda_{ij}} \tag{20}$$

$$\lambda_{0j} = \sum_{i \in \text{Inden}(1)} \lambda_{ij} \tag{21}$$

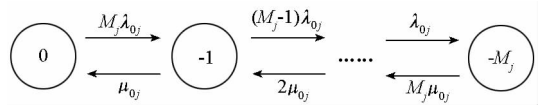


图4 装备的生灭过程模型

Fig.4 Birth and death process model for equipment

则现场站点 $j(j \in \text{Echelon}(H))$ 的装备可用度为:

$$A_j = 1 - EBO_{0j}/M_j = 1 - \left(\sum_{k=1}^{M_j} kP_{-k}^{0j} \right) / M_j \tag{22}$$

从而整个保障系统的可用度为:

$$A = \sum_{j \in \text{Echelon}(H)} M_j A_j / \sum_{j \in \text{Echelon}(H)} M_j \tag{23}$$

3 库存优化模型与算法

对于长期稳态下的备件配置问题,建立优化模型时一般以规定的保障效能指标(如装备可用度、备件满足率等)为约束条件,以寻求整个保障体系内备件总费用的最低值。本文选取部队常用的可用度作为保障效能指标,数学模型如下^[16]:

$$\begin{cases} \min & \sum_j \sum_i \tau_i S_{ij} \\ \text{s. t.} & A \geq A_0 \end{cases} \tag{24}$$

边际分析法是求解上述模型的常用方法,其操作简单,运算效率高,迭代过程中通过对增加单项备件导致的成本增量和收益增量进行综合分析以选择效益最高的备件项目。具体步骤如下:

步骤1:构造 I 行 J 列备件方案矩阵 S ,并令 $S = 0$;

步骤2:计算站点 j 第 i 项备件的边际效应值,即:

$$\delta(i,j) = \frac{A(S + \text{ones}(i,j)) - A(S)}{\tau_i} \tag{25}$$

其中,ones(i,j)表示第 i 行第 j 列元素为1,其余元素全为0的矩阵; $A(S)$ 表示备件方案 S 下的装备可用度;

步骤3:将 $\delta(i,j)$ 最大值所对应的站点 j_{\max} 上的备件 i_{\max} 库存量加1,由此得到新的库存量矩阵 $S = S + \text{ones}(i_{\max}, j_{\max})$;

步骤4:计算在新库存方案 S 下的装备可用度 A ,并与规定的可用度指标 A_0 比较,如果 $A \geq A_0$,算法结束,此时的 S 即为最优库存配置结果;反之则转入步骤2继续迭代。

4 仿真算法设计

采用 MATLAB 平台构造仿真模型对多等级多层级下的备件方案进行评估。仿真采用事件调度法,以故障发生作为原始驱动事件,由此引发后续一系列的维修、更换活动。由于故障发生后是否有备件更换取决于之前的故障维修情况,因此,采用与实际维修保障活动相反的事件顺序进行建模。对于多层级装备,建模顺序为:底层故障件的维修→底层备件数量更新→底层故障件的更换/上层故障件的维修→上层备件数量更新→上层故障件的更换→……直至顶层故障 LRU 的更换。对于多等级保障组织,如果故障件最终在顶层站点修复成功,建模顺序为:顶层站点故障件的维修→顶层站点备件数量更新→顶层站点备件下拨/下级站点备件数量更新→下级站点备件下拨→……直至底层站点备件数量更新和故障件的更换。

为了便于描述,以装备包含 LRU、SRU,保障体系为三等级结构为例,对仿真模型进行说明。其主流程如图5所示,包括库存信息模块、故障

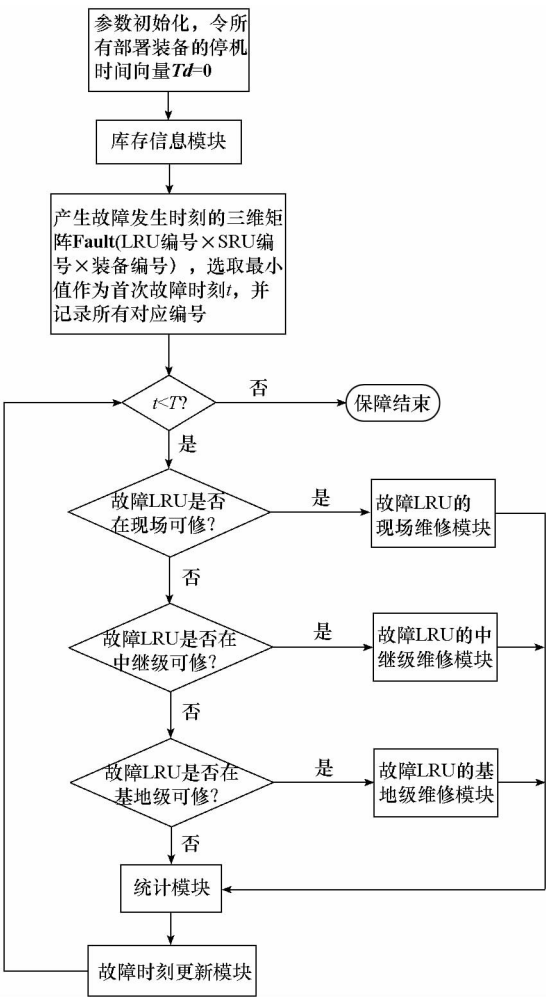


图5 单次仿真主流程

Fig.5 Main simulation process of single time

LRU 的现场维修模块、故障 LRU 的中继级维修模块、故障 LRU 的基地级维修模块、统计模块和故障时刻更新模块。为了保证模型进入稳态和仿真结果的稳定性,保障周期 T 和仿真次数取值应足够大。当一次仿真结束后, Td 中储存了每个装备的累积停机时间 $Td_{sys,i}$, 则各装备可用度为 $A_{sys,i} = 1 - Td_{sys,i}/T$, 现场站点 j 的可用度为站点所部署装备可用度的平均值为:

$$A_j = \sum_i A_{sys,i}/M_j$$

(26)

整个保障系统的可用度计算方法同式(23)。

5 算例分析

5.1 案例想定与结果分析

假设由图 2 所示的三级保障体系,对飞机上的某型装备开展维修供应保障,现要求制定各级站点的备件配置方案使整个保障系统的装备可用

度值不低于 0.95。装备的组成结构如图 1 所示。装备在三个现场站点的配备数量分别为 2,2,3, 平均每天工作时间均为 12 h。中继级站点至三个现场站点的补给运输时间分别为 4 d、5 d、6 d,基地级站点至两个中继级站点的补给运输时间分别为 8 d、10 d。备件及站点的相关可靠性、维修性等参数如表 1 所示。

边际算法历经 43 次迭代,得到最优备件配置方案,如表 2 所示,迭代曲线如图 6 所示,此时的总购置费用 $C = 547.5$ 万元,整个保障系统的装备可用度 $A = 0.9506$ 。从该备件方案可以看出:现场站点配备的 LRU 数量较多而 SRU 数量较少;基地级站点配备的 LRU 数量较少而 SRU 数量较多;中继级站点配备的 LRU、SRU 数量则居于现场站点和基地站点之间。由此可知,模型所得备件方案符合多级保障模式下的备件配置规律。

表 1 备件输入参数
Tab.1 Input parameters of the spare parts

Item	$MTBF_i/h$	Z_i	$\tau_i/\text{万元}$	T_{ij1}/d	T_{ij2}/d	T_{ij3}/d	T_{in1}/d	T_{in2}/d	T_{i10}/d	r_{ij1}	r_{ij2}	r_{ij3}	r_{in1}	r_{in2}	r_{i10}
LRU ₁	682.7	1	15	5	5	5	8	8	10	0.45	0.45	0.45	0.65	0.65	1
LRU ₂	333.3	1	20	8	8	8	10	10	12	0.60	0.60	0.60	0.70	0.70	1
LRU ₃	600	1	30	7	7	7	9	9	11	0.55	0.55	0.55	0.75	0.75	1
SRU ₁₁	2200	2	4	4	4	4	7	7	10	0.30	0.30	0.30	0.55	0.55	1
SRU ₁₂	1800	1	7	5	5	5	6	6	9	0.40	0.40	0.40	0.60	0.60	1
SRU ₂₁	1500	2	5.5	7	7	7	10	10	12	0.35	0.35	0.35	0.50	0.50	1
SRU ₂₂	1200	2	4.5	6	6	6	9	9	10	0.45	0.45	0.45	0.65	0.65	1
SRU ₃₁	2000	2	10	3	3	3	5	5	8	0.20	0.20	0.20	0.45	0.45	1
SRU ₃₂	3000	2	5	5	5	5	7	7	9	0.25	0.25	0.25	0.50	0.50	1

表 2 最优配置方案
Tab.2 Optimization result of spare parts configuration

Item	b_0	n_1	n_2	j_1	j_2	j_3
LRU ₁	0	1	0	1	1	3
LRU ₂	0	1	1	2	2	3
LRU ₃	0	1	0	1	1	2
SRU ₁₁	1	1	1	0	1	0
SRU ₁₂	1	0	0	0	0	0
SRU ₂₁	1	1	1	0	1	1
SRU ₂₂	1	1	1	1	1	1
SRU ₃₁	1	1	0	0	0	1
SRU ₃₂	1	1	1	0	0	1

将最优备件方案输入仿真模型,得到各现场站点和整个保障系统的装备可用度结果如表 3 所示,表中加入了解析模型的结果进行对比。从表

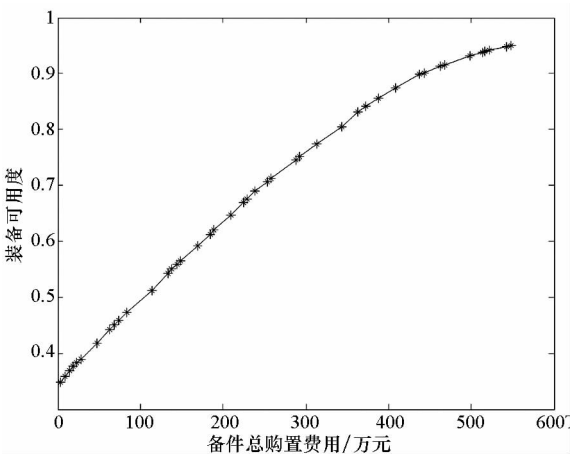


图 6 边际算法迭代曲线
Fig.6 Iterative curve by using marginal algorithm

中可以看出,解析模型的计算结果比仿真模型略微偏高,但误差均在合理的范围内,产生误差的原因是在多级保障模式下,备件到达率并没有严格

地满足生灭过程的使用条件,因此利用生灭过程建模是一种近似方法,但从算例结果看,该近似方法合理有效,所得的备件优化方案基本能满足保障要求。另外,通过大量算例表明,本节算例中解析模型的结果偏高并非个例,而是一种普遍现象(详见 5.2 节)。

表 3 解析模型与仿真模型的可用度结果对比

Tab. 3 Comparison of availability between analytic model and simulation model

可用度	A_{j1}	A_{j2}	A_{j3}	A
解析模型	0.936 6	0.938 7	0.967 9	0.950 6
仿真模型	0.914 0	0.918 1	0.945 0	0.925 8
相对误差	2.47%	2.24%	2.42%	2.68%

为了进一步提高保障的精确性,可采取解析与仿真相结合的方法对优化方案进行调整。由于解析模型对可用度的评估结果偏高,调整方法为适当增加边际算法的可用度指标值,使边际算法在初次所得优化方案的基础上继续迭代计算从而得到新的优化方案,并利用仿真模型检验其是否达到原始可用度指标要求,若不满足还可继续调整迭代直到满足为止。这样可通过若干次微调得到更为精准和满意的结果。例如,算例中优化方案的仿真可用度评估结果为 0.928 5,与 0.95 的指标要求仍存在近 0.02 的差距,将边际算法中的可用度指标值增加 0.02(设为 0.97),使边际算法在原备件方案基础上历经 6 次迭代,得到新的优化方案如表 4 所示。经过仿真模型验证,该方案的“真实”可用度结果为 0.953 5,达到了 0.95 的指标要求,则该方案为调整后的最终备件方案,此时的总购置费增加至 623 万元。

表 4 调整后的最优配置方案

Tab. 4 Optimization result of spare parts configuration after adjustment

Item	b_0	n_1	n_2	j_1	j_2	j_3
LRU ₁	0	1	0	2	2	3
LRU ₂	0	1	1	2	2	3
LRU ₃	0	1	0	1	1	3
SRU ₁₁	1	1	1	0	1	0
SRU ₁₂	1	0	0	0	0	0
SRU ₂₁	1	1	1	1	1	1
SRU ₂₂	1	1	1	1	1	1
SRU ₃₁	1	1	0	0	0	1
SRU ₃₂	1	1	1	1	1	1

5.2 VMETRIC 软件的验证与对比分析

VMETRIC^[16]是美国 TFD 公司开发的多等级多层级备件优化工具,该软件以 VARI-METRIC 模型为基本内核,可以实现对多个维修供应等级、多个层次备件配置方案的优化,曾多次用于美国各军兵种的装备采办、保障性分析、维修保障方案决策等,具有一定的权威性。本节利用仿真模型对 VMETRIC 软件和本文解析模型进行综合对比分析,对 VMETRIC 软件的准确性和精度进行验证和讨论。

5.2.1 优化性能对比分析

以 5.1 节中的案例为例,在 VMETRIC 软件中建立相关模型和想定。将可用度上限设为 0.98,分别利用本文解析模型与 VMETRIC 软件进行计算,得到的可用度迭代曲线如图 7 所示。从图中可以看出,在可用度不高的迭代前期,两条曲线的差异较大。特别地,在迭代起始点,当备件费用为 0 时(不配备件),本文模型的可用度计算结果为 0.34,而 VMETRIC 软件的结果为 0.02。但在迭代后期,当备件费用超过 400 万元、可用度大于 0.8 后,两者结果逐渐接近,最终在费用为 700 万元左右时,迭代相继结束,达到 0.98 的规定指标值。

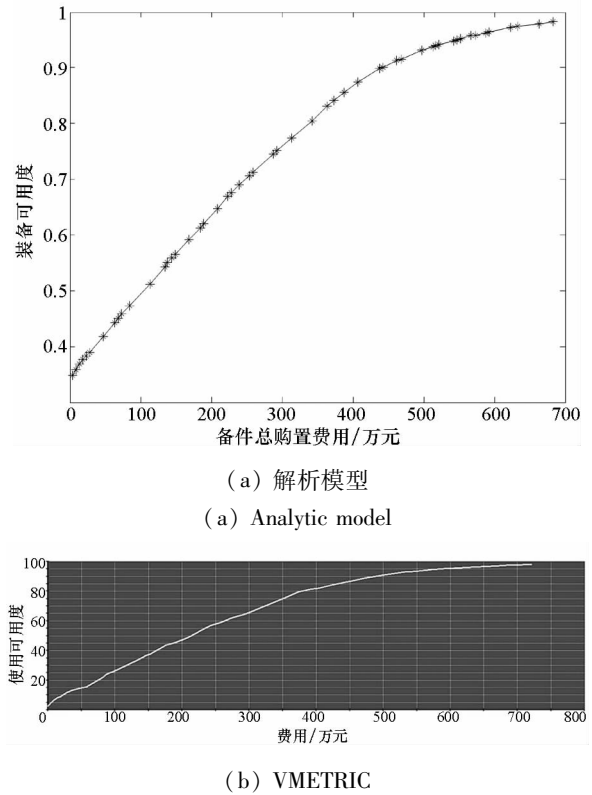


图 7 解析模型与 VMETRIC 软件的可用度迭代曲线对比

Fig. 7 Comparison of iterative curve for availability between analytic model and VMETRIC

由于与 VMETRIC 软件结果存在明显的差异,分别选取两条曲线上可用度迭代值达到 0.4, 0.6,0.9,0.98 的备件优化方案输入仿真模型进行评估和验证,所得结果如表 5~8 所示。表中加入了利用仿真模型对解析模型修正后的结果(解析-仿真模型)。从表中结果可以看出:

表 5 $A_0=0.4$ 时各模型方法对比
Tab.5 Comparison of results among different methods when $A_0=0.4$

模型	解析模型	解析-仿真模型	VMETRIC
计算结果	0.417 7	0.403 0	0.409 4
仿真评估	0.403 0	—	0.561 3
相对误差	3.65%	—	27.06%
总费用/万元	48	48	167

表 6 $A_0=0.6$ 时各模型结果对比
Tab.6 Comparison of results among different methods when $A_0=0.6$

模型	解析模型	解析-仿真模型	VMETRIC
计算结果	0.612 6	0.600 1	0.619 8
仿真评估	0.590 0	—	0.693 3
相对误差	3.83%	—	10.60%
总费用/万元	184	189	274

表 7 $A_0=0.9$ 时各模型方法对比
Tab.7 Comparison of results among different methods when $A_0=0.9$

模型	解析模型	解析-仿真模型	VMETRIC
计算结果	0.901 1	0.903 8	0.910 1
仿真评估	0.878 8	—	0.893 3
相对误差	2.54%	—	1.88%
总费用/万元	442.5	498	500

表 8 $A_0=0.98$ 时各模型方法对比
Tab.8 Comparison of results among different methods when $A_0=0.98$

模型	解析模型	解析-仿真模型	VMETRIC
计算结果	0.982 3	0.980 1	0.981 7
仿真评估	0.970 3	—	0.970 6
相对误差	1.24%	—	1.14%
总费用/万元	683	737	723

1)在可用度较低时(0.4、0.6),VMETRIC 软件与仿真结果差异很大,最大相对误差为 27.06%,具有极强的保守性;而解析模型的相对误差较小,最大值为 3.83%;

2)在可用度较高时(0.9、0.98),VMETRIC 软件 and 解析模型的误差均在合理范围内,但 VMETRIC 软件的误差无疑更小,优化性能强于解析模型。

3)采用仿真模型对解析模型进行修正的方法可以得到比 VMETRIC 软件更符合指标要求的结果,且能找到费效比更优的备件方案。如当 $A_0=0.9$ 时,本文模型计算得到的备件方案以 498 万元的费用使可用度达到 0.903 8,而 VMETRIC 的备件方案在 500 万元更多耗费的情况下其可用度仅为 0.893 3。

由此可以看出,VMETRIC 软件在高可用度时(0.8 以上)计算结果精确,而在低可用度下计算结果明显偏低,并不能对备件方案进行准确的定量评估。究其原因,是 VMETRIC 软件的核心算法——VARI-METRIC 模型导致的,因为 VARI-METRIC 模型算法中存在两个关键近似:一是用各 LRU 可用度的乘积去计算装备可用度,将导致结果偏低;二是对需求率的计算没有考虑装备停机时的影响(装备钝化)使需求率的估计比实际偏高,也将导致可用度结果偏低。以上两个因素在可用度较高时,影响微乎其微,而当可用度逐渐降低时,影响会越来越明显。这与算例结果完全吻合。尽管如此,但从实际使用的角度,由于保障人员关心的一般是高可用度指标下的备件方案,在此情况下,VMETRIC 不失为一款实用、优化性能优良、计算结果精确的软件。而本文解析模型在此案例中不论对于低可用度还是高可用度,其计算结果与仿真评估值相差均在 4% 以内,因此可以对任意备件方案下的可用度进行较准确的评估。

5.2.2 参数对模型精度的影响分析

为了更全面地了解模型的适应性,本节通过改变典型的输入参数值,分析它们对 VMETRIC 软件和本文解析模型精度的影响。

1)维修时间对模型精度的影响分析

仅以 LRU 在现场站点的维修时间为例,在 5.1 节的案例想定中保持其他参数不变,将所有现场站点 LRU 的维修时间 T_{LRU} 分别均设为 5 d、10 d、15 d、20 d、25 d、30 d,在表 4 所示的备件方案下利用 VMETRIC 软件、解析模型、仿真模型对可用度进行评估计算,所得结果如表 9 所示。从表中可以看出,在一定区间内随着 LRU 维修时间的增加,

VMETRIC 软件误差逐渐增大,而解析模型误差反而逐渐减小。这是因为解析模型是近似利用生灭过程建模,由于仅假设维修时间服从指数分布,因此实际备件到达时间并不服从指数分布。而当维修时间较长,维修时间对备件到达时间的影响会更大,使备件到达时间更贴近指数分布,从而使生灭过程建模产生的误差更小。其他站点的其他部件其维修时间的影响与此相同,不再举例。

表 9 不同维修时间下各模型的可用度结果对比
Tab.9 Comparison of availability among models with different maintenance time

模型	解析模型	仿真模型	VMETRIC
$T_{LRU} = 5\text{ d}$	0.979 2	0.963 5	0.973 9
$T_{LRU} = 10\text{ d}$	0.960 1	0.942 1	0.949 4
$T_{LRU} = 15\text{ d}$	0.934 9	0.920 4	0.914 5
$T_{LRU} = 20\text{ d}$	0.904 7	0.891 7	0.869 2
$T_{LRU} = 25\text{ d}$	0.871 0	0.864 6	0.814 0
$T_{LRU} = 30\text{ d}$	0.835 6	0.835 2	0.749 9

2) 运输时间对模型精度的影响分析

仅以中继级站点到现场站点的运输时间为例,在 5.1 节的案例想定中保持其他参数不变,将所有中继级站点至现场站点的运输时间 O 分别均设为 5 d、10 d、15 d、20 d、25 d、30 d,在表 4 所示的备件方案下利用 VMETRIC 软件、解析模型、仿真模型对可用度进行评估计算,所得结果如表 10 所示。从表中可以看出,在一定区间范围内当运输时间增大时,VMETRIC 软件误差较小,而解析模型误差越来越大。这是因为运输时间也是备件到达时间的重要决定部分,由于假设其为常数,并非指数分布,因此运输时间越大,运输时间对备件到达时间的影响会更大,使备件到达时间与指数分布的贴近度较低,从而使生灭过程建模产生的误差更大。其他站点运输时间的影响与此相同,不再举例。

表 10 不同运输时间下各模型的可用度结果对比
Tab.10 Comparison of availability among models with different transportation time

模型	解析模型	仿真模型	VMETRIC
$O = 5\text{ d}$	0.969 8	0.952 9	0.966 1
$O = 10\text{ d}$	0.940 2	0.916 6	0.935 4
$O = 15\text{ d}$	0.907 0	0.873 3	0.893 0
$O = 20\text{ d}$	0.868 1	0.826 5	0.838 7
$O = 25\text{ d}$	0.833 9	0.781 3	0.773 5
$O = 30\text{ d}$	0.803 6	0.743 6	0.698 5

6 结论

1) 无论是本文提出的解析模型还是 VMETRIC 软件,在建模过程中采取了一些近似处理手段,均无法保证在任何情况下都能得到精确甚至是正确的结果。具体表现为:VMETRIC 在高可用度时(0.8 以上)计算结果精确,而在低可用度或维修时间过长时计算结果明显偏低,并不能对备件方案进行准确的定量评估;本文解析模型在运输时间相对较短(低于各部件平均故障间隔时间的最小值)或维修时间相对较长时,基本能保证结果的准确性(误差在 4% 以内),但当运输时间过长时,也无法对备件方案进行准确的定量评估。

2) 虽然仿真模型结果更精确可信,但面对大批量备件的优化计算时,耗时将不可估量。因此,首先通过解析模型快速计算初始优化方案,再利用仿真模型对初始优化方案实施评估和调整,从而得到的最终优化方案(本文解析-仿真模型)无疑是更好的选择。

参考文献(References)

[1] Ruan M Z, Luo Y, Li H. Configuration model of partial repairable spares under batching ordering policy on inventory state[J]. Chinese Journal of Aeronautics, 2014, 27(3): 558-567.

[2] Sherbrooke C C. METRIC: a multi-echelon technique for recoverable item control[J]. Operations Research, 1968, 16(1): 122-141.

[3] Muckstadt J. A model for a multi-item, multi-echelon, multi-indenture inventory system[J]. Management Science, 1973, 20(4): 472-481.

[4] Sherbrooke C C. VARI-METRIC: improved approximations for multi-indenture multi-echelon availability models[J]. Operations Research, 1986, 34(2): 311-319.

[5] Sleptchenko A, Van Der Heijden M C, Van Harten A. Using repair priorities to reduce stock investment in spare part networks[J]. European Journal of Operational Research, 2005, 163(3): 733-750.

[6] Adan I, Sleptchenko A, Van Houtum G J. Reducing costs of spare parts supply system via static priorities[J]. Asia-Pacific Journal of Operational Research, 2009, 26(4): 559-585.

[7] Van Der Heijden M C, Alvarez E M, Schutten J M J. Inventory reduction in spare part networks by selective throughput time reduction[J]. International Journal of Production Economics, 2013, 143(8): 509-517.

[8] Salman S, Cassady C R, Pohl E A, et al. Evaluating the impact of cannibalization on fleet performance[J]. Quality and Reliability Engineering International, 2007, 23(4): 445-457.

[9] 李羚伟, 张建军, 张涛, 等. 面向任务的拼修策略问题及其求解算法[J]. 系统工程理论与实践, 2009, 29(7): 97-104.

LI Lingwei, ZHANG Jianjun, ZHANG Tao, et al. Mission oriented cannibalization policy problem and its solving algorithm [J]. Systems Engineering—Theory & Practice, 2009, 29(7): 97–104. (in Chinese)

[10] 罗伟, 阮旻智, 李庆民. 多级维修供应下不完全串件系统可用度评估[J]. 系统工程与电子技术, 2012, 34(6): 1182–1186.

LUO Yi, RUAN Minzhi, LI Qingmin. Evaluation of availability for incomplete cannibalization system under multi-echelon maintenance supply [J]. Systems Engineering and Electronics, 2012, 34(6): 1182–1186. (in Chinese)

[11] Olsson F. An inventory model with unidirectional lateral transshipments [J]. European Journal of Operations Research, 2010, 200(3): 725–732.

[12] Tiacci L, Saetta S. Reducing the mean supply delay of spare parts using lateral transshipments policies [J]. International Journal of Production Economics, 2011, 133(1): 182–191.

[13] 刘任洋, 李庆民, 李华. 基于横向转运策略的可修件三级库存优化模型[J]. 航空学报, 2014, 35(12): 3341–3349.

LIU Renyang, LI Qingmin, LI Hua. Optimal model of three-echelon inventory for repairable spare parts with lateral transshipments[J]. Acta Aeronautica et Astronautica Sinica, 2014, 35(12): 3341–3349. (in Chinese)

[14] 刘任洋, 黎放, 李庆民, 等. 基于横向转运策略的不完全修复件库存配置与订购模型[J]. 航空学报, 2015, 35(6): 1964–1974.

LIU Renyang, LI Fang, LI Qingmin, et al. Modeling for inventory distribution and ordering of imperfect repairable spare parts with lateral transshipments[J]. Acta Aeronautica et Astronautica Sinica, 2015, 35(6): 1964–1974. (in Chinese)

[15] 任敏, 陈全庆, 沈震, 等. 备件供应学[M]. 北京: 国防工业出版社, 2013.

REN Min, CHEN Quanqing, SHEN Zhen, et al. The supplying of the spare parts[M]. Beijing: National Defense Industry Press, 2013. (in Chinese)

[16] Sherbrooke C C. Optimal inventory modeling of system: multi-echelon techniques[M]. 2nd ed. Boston, US: Artech House, 2004.

两段探测目标的传感器任务调度问题 0-1 规划模型及算法*

李建平¹, 张 晗¹, 罗 永¹, 朱 承², 何文涛²

(1. 国防科技大学 理学院, 湖南 长沙 410073; 2. 国防科技大学 信息系统工程重点实验室, 湖南 长沙 410073)

摘 要:为解决指挥系统控制中的调度困难,研究了一类特殊的传感器资源调度问题。主要分析了跟踪目标的探测次数、时间间隔和传感器资源等约束条件。用跟踪目标的重要程度之和作为目标函数,建立了一个 0-1 规划的数学模型,再利用变换将其转化为 0-1 线性整数规划模型。利用割平面法求解得出最优调度策略,其能在工作量饱和的情况下合理调度传感器资源。为提高求解速度,提出了对应的模拟退火算法。通过对一些不同规模实例的求解,在资源利用率和算法的求解速度等指标上,与割平面法及遗传算法进行对比分析,验证了模型的有效性和模拟退火算法求解的高效性。

关键词:传感器;任务调度;0-1 规划;模拟退火算法;遗传算法

中国分类号:O221. 4 文献标志码:A 文章编号:1001-2486(2017)03-121-09

The 0-1 programming model and algorithm for the problem of sensor task scheduling for double detection

LI Jianping¹, ZHANG Han¹, LUO Yong¹, ZHU Cheng², HE Wentao²

(1. College of Science, National University of Defense Technology, Changsha 410073, China;
2. The Key Laboratory of Information System Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: In order to resolve the scheduling difficulty in the control of command system, the resource scheduling problem of a special sensor was studied and the constraint conditions including detected times, the interval between two detections, and the resource restrict of sensor were analyzed. A 0-1 programming model was established and transformed to a 0-1 liner integer model whose objective function is the sum of the importance degree of tracking targets. The optimal solution which can reasonably schedule sensor resource when the workload is saturated was obtained by using the cutting plane algorithm. A corresponding simulated annealing algorithm was proposed to improve the speed of solving and was used to solve some examples. Compared with the cutting plane algorithm and the genetic algorithm in terms of resource utilization and the speed of solving, the validity of the proposed model and the high efficiency of the simulated annealing algorithm were proved.

Key words: sensor; task scheduling; 0-1 programming; simulated annealing algorithm; genetic algorithm

传感器的任务调度是根据一定优化准则,在确定的时间区间内,为传感器网络中的传感器资源安排跟踪任务,以满足对多个移动目标的跟踪要求,从而达到某项或某些指标最优。传感器资源调度能够有效地提高资源利用效率。在军事问题中,合理的传感器调度,例如雷达,可以提高军方的防御能力。在研究的问题中,传感器对目标进行跟踪时需要先后进行两段探测,每段都有最短时长,若一个目标两段探测由不同的传感器完成,则两段探测之间有最短时间间隔。不同移动方向的目标有不同的重要程度,由于传感器资源的限制,当目标过多时,必须有选择地跟踪重要目标,由此产生了这种两段探测目标的传感器资源

调度问题。

目前,许多传感器资源调度问题主要是研究其启发式调度策略。文献[1-3]对基于工作方式优先级的驻留调度算法进行了改进,文献[4]首先提出了驻留时间窗的概念,其中时间窗^[5]是指雷达的实际执行时间在期望发射时间前后能移动的有效范围。驻留时间窗这一概念的提出增大了雷达任务调度的灵活性,原本冲突的任务有了被重新调度的可能。脉冲交错技术^[6]的使用可以提高雷达的时间利用率。Shih 等先后提出了最小空闲模版优先^[7]和基于模版的综合周期^[8]调度算法,这两种算法均为离线模版算法。Gopalakrishnan 等还提出了在线设计模版^[9]的想

* 收稿日期:2016-05-05
基金项目:国家自然科学基金资助项目(61273322)
作者简介:李建平(1965—),男,湖南涟源人,教授,博士,硕士生导师,Email:jpli_1@163.com

法。自适应调度算法^[10]是近年的研究热门。然而,对于两段探测目标的传感器资源调度问题仍然研究较少,目前文献已有的研究与实际问题应用仍有一定的差距。文献[11]提出了旋转情况下波束驻留任务的模型,并给出了一种旋转相控阵雷达的任务调度启发式算法。遗传算法在求解调度问题方面有广泛的应用。文献[12]给出了遗传算法对车间调度问题的求解方法;文献[13]提出一种新的改进自适应遗传算法,能够在优化过程中自动给出比较合适的交叉概率和变异概率,使算法在保持群体多样性的同时提高了速度;文献[14]等提出一种双层子代产生模式的改进遗传算法应用于车间调度问题,以使子代更好地继承父代的优良特征。

资源分配问题大多数归结为一个整数规划模型的求解。基于此事实,本文选用 0-1 变量对传感器调度问题进行建模和分析。

1 问题描述

当传感器探测目标时,每一个传感器都具有一定的探测范围,当该传感器的探测范围内有目标出现时,传感器可在极短时间内计算出目标移动的轨迹。即当目标进入探测区域内其任意时刻的位置视作已知。朝不同方向移动的目标有不同的目标重要程度,这是由目标本身决定的,传感器必须有选择地跟踪重要程度大的目标。

在传感器资源调度问题中,成功跟踪一个目标是指该目标在飞行过程中先后两次被(同一个或不同的)传感器探测到,并且每次探测的持续时间必须大于一个给定的时间,称这个给定的时间为最短探测时间。同时对某一目标的先后两次探测时间也需要有一定的时间间隔。两段探测可以由不同的传感器进行,但一段探测只能由一个传感器进行。若只探测一段或某段探测时间长度不够,则视为跟踪失败。虽然传感器可以同时探测多个目标,但其存在最大工作容量和探测范围的限制。最大工作容量是指传感器能同时探测的最大目标数量。

传感器资源调度问题的目标是寻找合适的调度策略使跟踪成功的目标重要程度之和最大。在这一问题中,需要决策的变量分别是对目标进行探测的传感器编号和两段探测开始的时刻。

2 问题分析与建模

2.1 模型基本假设

在工作饱和情况下,即目标数量超过最大工

作容量时,需要对多个目标有选择地进行跟踪。为避免重复,统一说明本文第 2 节中出现的 i 的取值范围是 $\{1, 2, \dots, n\}$, j 的取值范围是 $\{1, 2, \dots, m\}$, t 的取值范围是 $\{1, 2, \dots, N\}$ 以及上标 k 的取值范围是 $\{1, 2\}$ 。 n 表示任务期间出现的目标的总个数, w_i 表示目标 i 的重要程度。在任务期间传感器位置和探测范围是不变的,用 m 表示传感器总个数, C_j 表示传感器 j 的工作容量, N 表示时间段划分个数。

对于目标 i ,需要传感器探测的两段时间区间分别为 $s_i^1 = (t_i^1, t_i^1 + \Delta t_i^1)$, $s_i^2 = (t_i^2, t_i^2 + \Delta t_i^2)$,对应的传感器编号记作 R_i^1 和 R_i^2 。目标可能在任意时刻进入或离开传感器的探测区域,在模型中引入进入-离开时刻矩阵来描述目标进入和离开对应传感器探测范围的时间。两段探测目标的传感器调度的目标是确定对于每个目标 i 的探测起始时间 t_i^1, t_i^2 以及对应的探测传感器编号 R_i^1 和 R_i^2 ,使得被成功跟踪的目标重要程度之和最大,并满足下述约束条件:

1) 对目标 i 的两段探测时长 Δt_i^1 和 Δt_i^2 不小于给定的时长,在饱和和工作情况下探测时长为最短时长;

2) 对目标 i 的第一段探测结束后经过一定时间间隔才能开始第二段探测,即 $t_i^1 + \Delta t_i^1 + \Delta t_i < t_i^2$;

3) 对目标 i 的第二段探测结束时刻不晚于指定时刻 Te_i^2 ,即 $t_i^2 + \Delta t_i^2 < Te_i^2$;

4) 对于任意时刻 t ,传感器 j 同时探测的目标数量不超过其最大工作容量 C_j 。

引入 0-1 变量 X_{jt}^k 作为决策变量,其中 $X_{jt}^k = 1$ 表示从 t 时刻开始传感器 j 对目标 i 进行第 k 段探测($k = 1, 2$), $X_{jt}^k = 0$ 表示从 t 时刻开始传感器 j 对目标 i 没有进行第 k 段探测。

2.2 目标函数分析

为方便起见,引入一个效用值变量 u_i ,表示目标 i 是否两段都被探测。若目标 i 的两段探测都被完成,则 u_i 为 1,否则为 0。这样,目标函数为各 u_i 的重要程度的加权和,即 $\mathbf{w}^T \mathbf{u}$ 。模型中,目标函数为成功跟踪的目标的重要程度之和,即跟踪效用值的重要程度加权和 $z = \mathbf{w}^T \mathbf{u}$,其中 $\mathbf{w} = (w_1, w_2, \dots, w_n)^T$, $\mathbf{u} = (u_1, u_2, \dots, u_n)^T$ 。

2.3 约束条件分析

在工作饱和情况下,对上述约束条件 1~4 进行进一步分析,得到如下 4 个约束条件。

2.3.1 约束条件 1: 两段探测的限制

由决策变量 X_{jt}^k 的含义可知,对于固定的 i 和

k, X_{ijt}^k 等于1的个数即为目标 i 第 k 段被探测的次数,记为 u_i^k ,即:

$$u_i^k = \sum_{j=1}^m \sum_{t=1}^N X_{ijt}^k, \quad i=1,2,\dots,n; k=1,2 \quad (1)$$

在工作饱和情况下,只对目标的一段进行探测或者对某一段探测多于一次都是对资源的浪费。即对于目标 i ,应满足 u_i^1 和 u_i^2 同时为0或者1的条件。引入 n 个0-1辅助变量 z_1, z_2, \dots, z_n 。令

$$\sum_{j=1}^m \sum_{t=1}^N X_{ijt}^1 + z_i = 1, \quad i=1,2,\dots,n \quad (2)$$

$$\sum_{j=1}^m \sum_{t=1}^N X_{ijt}^2 + z_i = 1, \quad i=1,2,\dots,n \quad (3)$$

则当 $z_i=1$ 时, $u_i^1 = u_i^2 = 0$; 当 $z_i=0$ 时, $u_i^1 = u_i^2 = 1$ 。因此根据 u_i 的定义有:

$$u_i = \frac{1}{2}(u_i^1 + u_i^2) \quad (4)$$

2.3.2 约束条件2:传感器的最大工作容量限制

当传感器开始探测目标时,在此后的一段时间 Δt_i^1 或 Δt_i^2 内,传感器必须持续探测目标才有效,在这段时间内传感器资源处于占用状态,必须保证每个时刻传感器 j 同时探测的目标数目不超过最大工作容量 C_j 。用符号 y_{ijt} 表示传感器 j 在时刻 t 对目标 i 的探测情况。 $y_{ijt}=1$ 表示 t 时刻传感器 j 正在探测目标 i 。记 $\mathbf{Y} = (y_{ijt})_{n \times m \times N}$ 表示由 y_{ijt} 构成的三维矩阵。

下面推导 \mathbf{Y} 与 X_{ijt}^k 的关系式,考虑 $X_{ijt}^k=1$ 表示从 t 时刻到 $t+\Delta t_i^k$ 时刻之前传感器的资源始终被目标 i 占用,即对所有满足 $t \leq s < t+\Delta t_i^k$ 的时刻 s 有 $y_{ijs}=1$ 。用一个 $n \times m \times N$ 的三维矩阵 \mathbf{I}_{ijt}^k 表示 $X_{ijt}^k=1$ 时的决策对资源的使用情况。当满足 $t \leq s < t+\Delta t_i^k$ 时, $\mathbf{I}_{ijt}^k[i, j, s] = 1$, 其余情况取0。每一个 X_{ijt}^k 都对应一个 \mathbf{I}_{ijt}^k , 共 $2 \times n \times m \times N$ 个。

每一个矩阵 \mathbf{I}_{ijt}^k 表示在决策 $X_{ijt}^k=1$ 下资源的使用,将所有 $X_{ijt}^k=1$ 对应的 \mathbf{I}_{ijt}^k 累加,即可得出 \mathbf{Y} :

$$\mathbf{Y} = \sum_{k=1}^2 \sum_{i=1}^n \sum_{j=1}^m \sum_{t=1}^N X_{ijt}^k \times \mathbf{I}_{ijt}^k \quad (5)$$

由 \mathbf{Y} 的含义知,对于固定的 j 和 t ,使 $y_{ijt}=1$ 的 i 的个数即为传感器 j 在 t 时刻的工作量,其应不大于传感器 j 的最大工作容量 C_j ,即:

$$\sum_{i=1}^n y_{ijt} \leq C_j, \quad j=1,2,\dots,m; t=1,2,\dots,N \quad (6)$$

2.3.3 约束条件3:分段探测的先后时间限制

成功跟踪一个目标需要两段探测并且要保持两段探测的先后关系和时间间隔,即在第一段探测完成之后的 Δt_i 时长之前都不会开始第二段探测。

考虑工作饱和情况下第一段探测和第二段探测的次数均不超过一次(式(2)~(3)),所以对目标 i ,任意时刻 t_1 之后开始第一段探测的次数与 t_1 时刻之前开始第一段探测的次数的差只能取0和 ± 1 。取值为1表示第一段探测是在 t_1 时刻或之后开始,取值为-1表示第一段探测是在 t_1 时刻之前开始,取值为0表示放弃跟踪目标 i 。即:

$$\sum_{j=1}^m \sum_{t \geq t_1}^N X_{ijt}^1 - \sum_{j=1}^m \sum_{t < t_1}^N X_{ijt}^1 = 0 \text{ 或 } \pm 1 \quad (7)$$

当式(7)取值为1时,表示目标 i 的第一段探测发生在 t_1 时刻之后,为了有足够的时间进行第二段探测也该有 $t_1 + \Delta t_i^1 + \Delta t_i < N - \Delta t_i^2$ 。若满足两段探测的条件,第二段探测至少在 $t_1 + \Delta t_i^1 + \Delta t_i$ 之后开始。在此之前不应该进行第二段探测的任务分配,即:

$$\sum_{j=1}^m \sum_{t < t_1 + \Delta t_i^1 + \Delta t_i}^N X_{ijt}^2 = 0 \quad (8)$$

因为第二段探测最多一次,显然有:

$$\sum_{j=1}^m \sum_{t < t_1 + \Delta t_i^1 + \Delta t_i}^N X_{ijt}^2 \leq u_i^2 \leq 1 \quad (9)$$

由式(7)和式(9)得:

$$\sum_{j=1}^m \left(\sum_{t \geq t_1}^N X_{ijt}^1 - \sum_{t < t_1}^N X_{ijt}^1 + \sum_{t < t_1 + \Delta t_i^1 + \Delta t_i}^N X_{ijt}^2 \right) \leq 1 \quad (10)$$

式(10)要对所有满足 $t_1 + \Delta t_i^1 + \Delta t_i < N - \Delta t_i^2$ 的 t_1 成立。

下证式(10)也是约束条件3的充分条件。假设式(10)成立,设传感器 j 在 t_1 时刻对目标 i 进行第一段探测,则:

$$\sum_{j=1}^m \sum_{t \geq t_1}^N X_{ijt}^1 - \sum_{j=1}^m \sum_{t < t_1}^N X_{ijt}^1 = 1 \quad (11)$$

由式(10)得:

$$\sum_{j=1}^m \sum_{t < t_1 + \Delta t_i^1 + \Delta t_i}^N X_{ijt}^2 = 0 \quad (12)$$

即 $t_1 + \Delta t_i^1 + \Delta t_i$ 时刻之前没有对目标 i 进行第二段探测,由此可知第二段探测在 $t_1 + \Delta t_i^1 + \Delta t_i$ 时刻之后,充分性得证。综上,式(10)是约束条件3的充要条件。

2.3.4 约束条件4:进入-离开时间限制

若开始探测的时间太靠后,则一段探测未完成目标就可能离开了探测区域,则需要重新选择

开始探测时刻或放弃跟踪。用 T_{ij}^1 表示目标 i 进入传感器 j 区域的时刻,用 T_{ij}^2 表示目标 i 离开传感器 j 区域的时刻。若目标 i 不进入传感器 j 扫描区域,取 $T_{ij}^1 = T_{ij}^2 = 0$ 。

1) 情况 1: $T_{ij}^1 = T_{ij}^2 = 0$, 即目标 i 不进入传感器 j 扫描区域, 令

$$X_{ijt}^1 = 0, \quad t = 1, 2, \cdots, N \tag{13}$$

$$X_{ijt}^2 = 0, \quad t = 1, 2, \cdots, N \tag{14}$$

2) 情况 2: ①当 $t + \Delta t_i^1 > T_{ij}^2 + 1$, 即目标 i 第一段探测的结束时间晚于离开传感器 j 扫描区域的时间, 令

$$X_{ijt}^1 = 0, \quad t > T_{ij}^2 - \Delta t_i^1 + 1 \tag{15}$$

②类似地, 当 $t + \Delta t_i^2 > T_{ij}^1 + 1$, 令

$$X_{ijt}^2 = 0, \quad t > T_{ij}^1 - \Delta t_i^2 + 1 \tag{16}$$

2.4 0-1 线性整数规划模型

为了求解方便,在整数规划中,一般习惯用列向量表示决策变量,用矩阵表示约束条件。将由 X_{ijt}^k, z_i 共 $2nmN + n$ 个变量构成的决策变量变换为一维列向量 \mathbf{v} 。将三种变量字典顺序排列: X_{ijt}^1 下标对应为 $(i-1)mN + (j-1)N + t$; X_{ijt}^2 下标对应为 $(n+i-1)mN + (j-1)N + t$; z_i 下标对应为 $2nmN + i$ 。这样每一个 0-1 变量都在 \mathbf{v} 中有固定的位置。

约束分为等式约束和不等式约束,分别是式(2)、式(3)、式(13)~(16)六组等式和式(6)、式(10)两组不等式。约束矩阵分为等式约束矩阵 \mathbf{Aeq} 和不等式约束矩阵 \mathbf{A} 。 \mathbf{Aeq} 的各行分别是式(2)、式(3)、式(13)~(16)六组等式中左端决策变量的系数按 \mathbf{v} 中决策变量的顺序排列的行向量。类似地,约束矩阵 \mathbf{A} 对应于式(6)、式(10)两组不等式。右端向量 \mathbf{beq} 和 \mathbf{b} 分别是 \mathbf{Aeq} 和 \mathbf{A} 对应约束等式和不等式的右侧常数项排列组成的列向量。目标函数为:

$$\mathbf{z} = \mathbf{w}^T \mathbf{u} = \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^2 \sum_{j=1}^m \sum_{t=1}^N w_i X_{ijt}^k \tag{17}$$

令 \mathbf{c} 表示 \mathbf{v} 中决策变量在目标函数中的系数排列成的列向量,于是,模型可以写成标准的线性 0-1 整数规划格式:

$$\begin{aligned} & \text{(IP) max } \mathbf{c}^T \mathbf{v} \\ & \text{s. t. } \begin{cases} \mathbf{A} \mathbf{v} \leq \mathbf{b}, \\ \mathbf{Aeq} \cdot \mathbf{v} = \mathbf{beq}, \end{cases} \\ & \mathbf{v} \in \{0, 1\}^r, r = 2nmN + n \end{aligned} \tag{18}$$

3 基于模拟退火的求解算法

目前,用来求解规模较大的优化和调度问题的常用方法是智能搜索算法,如遗传算法、粒子群

算法、禁忌搜索算法、模拟退火算法等。这些算法能够在可接受的时间内得出近似最优解。同其他的智能搜索算法相比,模拟退火算法收敛性较好,理论上可以证明以概率 1 收敛到最优解,并且具有描述简单、使用灵活、运行效率高和较少受到初始解影响等优点,因此得到较多关注并应用于一些优化问题,取得了较好的效果。

3.1 算法的要素

模拟退火算法是一种模拟固体退火过程的算法。使用模拟退火算法时需给出初始温度、终止温度、降温函数、能量函数、解的邻域和移动等算法要素。在本例中降温函数选用 $T_k = T_0 / [\log(1 + k)]$, $k \in \mathbb{N}^+$ 。能量函数是模拟退火算法中表示当前解优劣的函数,能量越低对应的解越接近于最优。模拟退火算法试图找到使能量函数最小的解。因为原问题中目标是最大化被跟踪目标的重要程度之和,所以本例中能量函数 f 用未被跟踪的目标的重要程度之和表示。

模型中一个解的邻域是由从这个解对应的调度策略中增减调度方式产生的所有可行调度策略构成的集合。解的移动分为无条件移动和有条件移动。设 i 为当前解, f 为对应能量函数, j 为当前解邻域中的某个解。若 $f(j) < f(i)$, 则当前解可以进行无条件移动。反之,若 $f(j) > f(i)$, 则当前解 i 以一定概率转移至解 j 。

3.2 算法步骤

类比退火过程,将能量 E 模拟为目标函数值 f , 温度 T 演化成控制参数 t , 即得到解组合优化问题的模拟退火算法:由初始解 i 和控制参数初值 T_0 开始,对当前解重复“产生当前解邻域中的解→计算目标函数差→接受或舍弃”的迭代,并逐步衰减 T 值,当 T 低于终止温度时算法终止,算法终止时的当前历史最优解即为近似最优解,这是基于蒙特卡洛迭代求解法的一种启发式随机搜索过程。下面是算法的步骤:

步骤 1: 初始化。任选初试解 i , 给定初始温度 T_0 和终止温度 T_f , 确定降温函数 ΔT 和循环次数 $n(T_k)$, 令迭代指标 $k = 0$, $T_k = T_0$ 。

步骤 2: 随机产生 i 的一个邻域解 j , 计算目标值增量 $\Delta f = f(j) - f(i)$ 。

步骤 3: 若 $\Delta f < 0$, 令 $i := j$, 转步骤 4; 否则, 产生随机数 $\xi \in (0, 1)$, 若 $\exp\left(\frac{\Delta f}{T_k}\right) > \xi$, 则令 $i := j$ 。

步骤 4: 若达到热平衡(内循环次数大于 $n(T_k)$), 转步骤 5; 否则转步骤 2。

步骤 5:按降温函数降低温度 $T_k, k:=k+1$, 若 $T_k < T_f$, 停止迭代; 否则转步骤 2。

4 模拟实验及结果分析

4.1 实例构建

为进行仿真实验,构建一个由 3 个传感器

对 24 个目标进行跟踪的资源分配问题实例。其中,表 1 描述了模型的基本参数;表 2 根据目标所在的探测区域不同将目标分成六组,并列出了每组目标的重要程度;表 3 给出了上述每一组目标在不同传感器探测区域的进入和离开时刻。

表 1 基本参数
Tab.1 Basic parameters

目标数目	传感器数目	时间段数	最大工作容量	第一段探测最短时间	第二段探测最短时间	时间间隔
$n=24$	$m=3$	$N=11$	$C_j=2$	$\Delta t_i^1=2\text{ s}$	$\Delta t_i^2=3\text{ s}$	$\Delta t_i=0\text{ s}$

表 2 目标重要程度
Tab.2 Weight of targets

目标编号	1	2	3	4	5	6	7	8	9	10	11	12
重要程度	1	2	3	4	1	2	3	4	1	2	3	4
目标编号	13	14	15	16	17	18	19	20	21	22	23	24
重要程度	1	2	3	4	1	2	3	4	1	2	3	4

表 3 目标进入和离开传感器探测区域的时刻表(t_{ij}^1, t_{ij}^2)
Tab.3 Entry time and leave time of targets (t_{ij}^1, t_{ij}^2)

	目标 1	目标 2	目标 3	目标 4	目标 5	目标 6	目标 7	目标 8
传感器 1	(1,4)	(1,4)	(1,4)	(1,4)	(1,11)	(1,11)	(1,11)	(1,11)
传感器 2	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
传感器 3	(1,11)	(1,11)	(1,11)	(1,11)	(0,0)	(0,0)	(0,0)	(0,0)
	目标 9	目标 10	目标 11	目标 12	目标 13	目标 14	目标 15	目标 16
传感器 1	(1,5)	(1,5)	(1,5)	(1,5)	(0,0)	(0,0)	(0,0)	(0,0)
传感器 2	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)
传感器 3	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
	目标 17	目标 18	目标 19	目标 20	目标 21	目标 22	目标 23	目标 24
传感器 1	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
传感器 2	(1,2)	(1,2)	(1,2)	(1,2)	(0,0)	(0,0)	(0,0)	(0,0)
传感器 3	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)	(1,11)

4.2 割平面法与模拟退火算法实验结果对比分析

利用上述实例数据进行了仿真模拟,用割平面法求解式(18)所示 0-1 线性整数规划模型。调用 MATLAB 软件包求解,得到传感器对目标的跟踪策略,如表 4 所示。

从表 4 看出,按最优策略一共成功跟踪了 13 个目标,跟踪目标重要程度之和为 44。在式(18)所示模型中,去掉 0-1 整数约束,得到一个松弛的线性规划模型,利用单纯形法求解

得到目标函数最优值上界为 44.4。注意前面提出的 0-1 整数规划模型其目标函数值是整数,这说明,表 4 的传感器调度策略是最优策略。结合表 3 也可以看出重要程度大的目标均被跟踪。表 5 记录了最优调度策略下各传感器对各目标的具体跟踪情况。

表 5 中相邻单元格表示传感器对一个目标在多个时间段内连续的探测。可以看出,只有传感器 1 在时刻 8 的资源没有被使用,其余时间所有

传感器处于最大工作状态;由于每个目标需要 5 个时间段进行跟踪,资源的调度已经非常充分,并且满足了跟踪需要有前后两次探测的要求;没有出现无效的跟踪,如只探测一段或探测时间少于最短时间的情况。

通过上面的分析验证了模型的正确性,说明

该模型准确地描述了传感器资源调度实际问题。下面再用模拟退火算法对该模型求解进行比较分析,将目标的跟踪情况记录于表 6。

从表 6 可以看出,最优策略下成功跟踪了 12 个目标,跟踪目标重要程度之和为 41。

表 7 记录了传感器的资源调度情况。

表 4 跟踪情况——割平面法
Tab.4 Tracking status—cutting plane algorithm

目标	1	2	3	4	5	6	7	8	9	10	11	12
跟踪	否	否	是	是	否	是	是	是	否	否	是	是
目标	13	14	15	16	17	18	19	20	21	22	23	24
跟踪	否	否	是	是	否	否	是	是	否	否	是	是

表 5 传感器任务分配——割平面法
Tab.5 Task allocation of sensors—cutting plane algorithm

传感器	各时间段探测的目标											
	1	2	3	4	5	6	7	8	9	10	11	
传感器 3	目标 23	目标 23	目标 19	目标 19	目标 19	目标 20	目标 20	目标 20	目标 3	目标 3	目标 3	
	目标 24	目标 24	目标 23	目标 23	目标 23	目标 24	目标 24	目标 24	目标 4	目标 4	目标 4	
传感器 2	目标 19	目标 19	目标 11	目标 11	目标 11	目标 12	目标 12	目标 12	目标 15	目标 15	目标 15	
	目标 20	目标 20	目标 12	目标 12	目标 15	目标 15	目标 16	目标 16	目标 16	目标 16	目标 16	
传感器 1	目标 6	目标 6	目标 3	目标 3	目标 6	目标 6	目标 6	未使用	目标 7	目标 7	目标 7	
	目标 11	目标 11	目标 4	目标 4	目标 8	目标 8	目标 7	目标 7	目标 8	目标 8	目标 8	

表 6 跟踪情况——模拟退火算法
Tab.6 Tracking status—simulated annealing algorithm

目标	1	2	3	4	5	6	7	8	9	10	11	12
跟踪	否	否	是	是	否	否	是	是	否	是	是	是
目标	13	14	15	16	17	18	19	20	21	22	23	24
跟踪	否	否	是	是	否	否	否	是	否	否	是	是

表 7 传感器任务分配——模拟退火算法
Tab.7 Tack allocation of sensors—simulated annealing algorithm

传感器	各时间段探测的目标											
	1	2	3	4	5	6	7	8	9	10	11	
传感器 3	目标 20	目标 20	目标 4	目标 4	目标 4	目标 3	目标 3	目标 3	目标 20	目标 20	目标 20	
	目标 24	目标 24	未使用	目标 19	目标 19	目标 19	目标 19	目标 19	目标 24	目标 24	目标 24	
传感器 2	目标 10	目标 10	目标 10	目标 10	目标 10	未使用	目标 15	目标 15	目标 15	目标 15	目标 15	
	目标 11	目标 11	目标 11	目标 11	目标 11	未使用	目标 16	目标 16	目标 16	目标 16	目标 16	
传感器 1	目标 4	目标 4	目标 3	目标 3	未使用	未使用	目标 7	目标 7	目标 7	目标 7	目标 7	
	目标 11	目标 11	目标 11	目标 11	目标 11	未使用	目标 8	目标 8	目标 8	目标 8	目标 8	

图1是两种调度结果的甘特图。设目标编号是 M_i ,甘特图中 M_i01 表示第一段探测, M_i02 表示第二段探测。从图1可以看出,割平面法的最优策略下只有1个资源被浪费,而由模拟退火算法得到的策略中增加了5个被浪费的资源。因此,由模拟退火算法得到的调度策略的资源使用效率相比于最优调度策略仍然较低,并且目标函数最优值上也有一定的差距。但是在大规模实际问题的求解时间上,启发式算法的优势得到了体现。事实上,大规模情况中割平面法往往在规定的时间内得不到可行解,但启发式算法求解的时间长度并没有随问题规模增大而显著增加,并且其还能够较快地得到一个较优的解。在实际问题中,跟踪的目标数量应在100个左右,从调度开始到结束时间在800 s左右。由于运算精度的限制,时间段划分则至少在400段以上,要求算法在

利用割平面法求得: 跟踪数目=13 最优值=44

3号传感器	2301	1902			2002			302			
	2401	2302			2402			402			
2号传感器	1901	1102			1202			1502			
	2001	1201		1501	1601			1602			
1号传感器	601	301			602			未使用		702	
	1101	401		801	701		802				
时间段	1	2	3	4	5	6	7	8	9	10	11

利用模拟退火算法求得: 跟踪数目=12 最优值=41

3号传感器	2001			402			302			2002		
	2401			未使用			1001			1002		
2号传感器	1001			1002			未使用			1501		
	1101			1102			未使用			1601		
1号传感器	401			301			未使用			701		
	1101			1102			未使用			801		
时间段	1	2	3	4	5	6	7	8	9	10	11	

图1 甘特图
Fig.1 Gantt chart

20 s 内计算出最佳调度策略,这对算法的求解速度有很高的要求。从算法运行时间的对比可以看出,虽然模拟退火算法求解的目标函数值与最优值有一定的差距,但是在求解时间上有巨大的优势,已经基本满足了实际问题的需要。

一次调度在某一时间段内的资源利用率是指该时间段内使用的传感器资源的数目与总传感器的资源数目之比(每个传感器 j 有 C_j 个资源)。资源利用率指标可以直观地反映出一次调度的效率,资源利用率越高,调度的效率越高。图2表示的是两种算法求得的调度结果在每一时间段内的资源利用率,即资源利用率曲线。

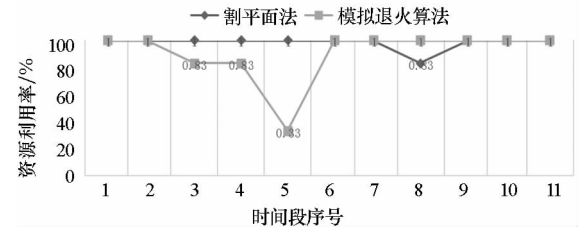


图2 资源利用率对比图

Fig.2 Comparison diagram of resource utilization

从图2可以看出割平面法产生的调度在资源使用上的优势,几乎每个时间段内全部的传感器资源都能被有效地调度起来,而对于模拟退火算法产生的调度,在时间段2至时间段6中出现了许多的资源浪费。

对于时间精度要求更高的模型,即时间段数划分更大时,启发式算法显示出求解速度快的优势。下面对表8中列出的基本参数的模型用割平面法和模拟退火算法求解,以对比两种算法的计算速度。

表8 基本参数

Tab.8 Basic parameters

目标数目	传感器数目	时间段数	最大工作容量	第一段探测最短时间	第二段探测最短时间	时间间隔
$n = 80$	$m = 5$	$N = 400$	$C_j = 8$	$\Delta t_i^1 = 80\text{ s}$	$\Delta t_i^2 = 100\text{ s}$	$\Delta t_i = 20\text{ s}$

在同一台PC上对表8中的模型求解,模拟退火算法速度是割平面法的10倍以上,并随着数据规模的增大,倍数会扩大。图3是使用模拟退火算法得到的资源利用率曲线。其中,最晚截止时间被控制在 $t = 340$,即在时间段340以后传感器不再跟踪目标。

为了仿真实验构建了一组不同规模的模拟实例,每个模拟实例都分别用模拟退火算法和割平面算法求解,运行时间、求解得到的最优值以及跟踪目标数于表9,其中SA表示模拟退火算法,CP

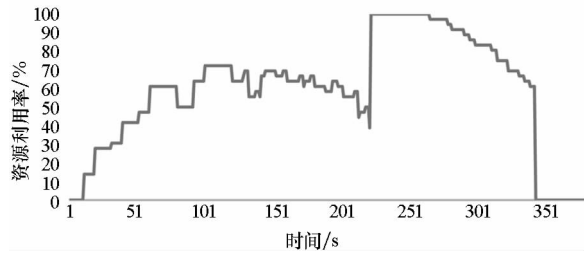


图3 模拟退火算法资源利用率图

Fig.3 Resource utilization of simulated annealing algorithm

表示割平面法。运行计算机的配置是 2.7 GHz Intel Core i5 处理器,8 G 内存。表格中案例 5 至案例 8 在利用割平面方法求解时内存超出限制,只能用启发式算法求解。

从表 9 可以看出模拟退火算法的运行时间随规模的增大并没有显著增加,也满足了规模为 100 个目标的实际问题中运行时间不超过 20 s 的限制。

表 9 运行时间和结果对比
Tab.9 Comparison of running time and results

规模		40 个目标		80 个目标		160 个目标		240 个目标	
案例标号		1	2	3	4	5	6	7	8
运行时间/s	SA	4	4	8	7	17	17	34	33
	CP	31	31	360	362	—	—	—	—
目标函数值	SA	32	44	78	64	70	86	138	119
	CP	34	47	72	74	—	—	—	—
跟踪目标数	SA	12	18	29	24	24	30	48	37
	CP	14	21	32	27	—	—	—	—

4.3 模拟退火算法与遗传算法实验结果对比分析

柔性作业车间调度问题常用遗传算法求解^[13-14],受此启发,将两段探测目标的传感器调度问题模型转化为一个特殊的柔性作业车间调度问题,用遗传算法求解,并与模拟退火算法对比。

传感器对一个目标的两段探测对应两道工序。将每个传感器 j 看作是 $C_j(j=1,2,\cdots,m)$ 个子机器的并集,每个子机器只能加工一个工件。目标被传感器跟踪时前后两段探测的探测时间对应于工件被机器加工时两道工序所需的加工时间。工件与其对应的目标有相同的重要程度。取遗传算法的适应值为在截止时间之前加工完成的工件的重要程度之和。

用模拟退火算法和遗传算法对四个模拟案例进行求解,结果见表 10。由表 10 可看出,遗传算法无论是在求解时间还是目标函数值方面效果均不如模拟退火算法。事实上,在遗传算法中,目标函数为在截止时间前被加工完成的的重要程

度之和,但是解码时整串编码都需要先转为一个调度,包括在截止时间后完成加工的目标,因此降低了算法的求解效率。

5 结论

双段探测目标的传感器资源调度问题是一种特殊的资源调度问题。本文以跟踪目标的探测次数、时间和传感器资源等为约束条件建立了 0-1 线性整数规划模型,从整数规划的角度对模型进行分析和求解,同时提出了对应的模拟退火算法。通过模拟退火算法与割平面法及遗传算法的对比发现模拟退火算法求解这类问题时在时间和效率上具有明显优势。对于小规模问题,割平面法能得到全局最优解,便于理论分析。提出的模型与调度算法在实际中能为双段探测目标的传感器资源调度问题提供一个可行的方案。如何借鉴车间调度问题的研究方法来解决双段探测目标的传感器资源调度问题值得进一步研究。

参考文献 (References)

[1] 张伯彦,蔡庆宇. 相控雷达阵的自适应调度和多目标数据预处理技术[J]. 电子学报,1997,25(9):1-5.
ZHANG Boyan, CAI Qingyu. Adaptive scheduling and multitarget data processing techniques of phased array radars[J]. Acta Electronica Sinica, 1997, 25(9): 1-5. (in Chinese)

[2] 李昊,于周秋. 基于优先级的相控阵测量雷达调度设计[J]. 现代雷达,2006,28(7):52-55.
LI Hao, YU Zhouqiu. Design of phased array instrumentation radar scheduling algorithm based on priority [J]. Modern Radar, 2006, 28(7): 52-55. (in Chinese)

[3] 孟宪福. 基于优先级的任务调度与负载均衡模型研究[J]. 小型微型计算机系统,2005,26(9):1601-1605.
MENG Xianfu. Study of task scheduling and load balancing

表 10 两种启发式算法的对比

Tab.10 Comparison of two heuristic algorithms

案例	案例	9	10	11	12
规模	目标数	240	160	160	80
模拟退火方法	跟踪目标数	44	29	24	12
	目标函数值	127	83	72	36
	运行时间/s	37	17	17	9
遗传算法	跟踪目标数	38	26	23	12
	目标函数值	86	62	54	30
	运行时间/s	116	68	66	34

models based on priority [J]. Mini-Micro Systems, 2005, 26(9): 1601-1605. (in Chinese)

[4] Huizing A G, Bloemen A A F. An efficient scheduling algorithm for a multifunction radar [C]// Proceedings of International Symposium on Phased Array Systems and Technology, 1996: 359-364.

[5] 何金新, 邱杰, 王国宏. 相控阵雷达事件调度中的时间窗研究[J]. 雷达科学与技术, 2010, 8(1): 80-86.
HE Jinxin, QIU Jie, WANG Guohong. Study on time window in multifunction phased array radar task scheduling [J]. Radar Science and Technology, 2010, 8(1): 80-86. (in Chinese)

[6] Farina A, Neri P. Multi-target interleaved tracking for phased-array[J]. IEE Proceedings, Part F—Communications, Radar and Signal Processing, 1980, 127(4): 312-318.

[7] Shih C S, Gopalakrishnan S, Ganti P. Scheduling real-time dwells using tasks with synthetic periods[C]//Proceedings of 24th IEEE International Real-Time Systems Symposium, 2003: 210-219.

[8] Shih C S, Ganti P, Gopalakrishnan S. Synthesizing task periods for dwells in multi-function phased array radars[C]// Proceedings of the IEEE Radar Conference, Philadel, 2004: 145-150.

[9] Gopalakrishnan S, Caccamo M, Shih C S. Finite-horizon scheduling of radar dwells with online template construction[C]// Proceedings of IEEE International Real-Time Systems Symposium, 2004: 23-33.

[10] 陈怡君, 罗迎, 张群, 等. 基于认知 ISAR 成像的相控阵雷达资源自适应调度算法[J]. 电子与信息学报, 2014, 36(7): 1566-1572.

CHEN Yijun, LUO Ying, ZHANG Qun, et al. Adaptive scheduling algorithm for phased array radar based on cognitive ISAR imaging [J]. Journal of Electronics & Information Technology, 2014, 36(7): 1566-1572. (in Chinese)

[11] 程小枫, 涂刚毅, 吴少鹏. 双波段旋转相控阵雷达任务调度算法[J]. 科学技术与工程, 2014, 14(23): 73-80.
CHENG Xiaofeng, TU Gangyi, WU Shaopeng. Task scheduling algorithm for dual-band rotating phased array radar[J]. Science Technology and Engineering, 2014, 14(23): 73-80. (in Chinese)

[12] 高亮, 张国辉, 王晓娟. 柔性作业车间调度智能算法及其应用[M]. 1 版. 武汉: 华中科技大学出版社, 2012.
GAO Liang, ZHANG Guohui, WANG Xiaojuan. Flexible shop scheduling intelligent algorithm and its application[M]. 1st ed. Wuhan: Huazhong University of Science and Technology, 2012. (in Chinese)

[13] 王万良, 吴启迪, 宋毅. 求解作业车间调度问题的改进自适应遗传算法[J]. 系统工程理论与实践, 2004, 24(2): 58-62.
WANG Wanliang, WU Qidi, SONG Yi. Modified adaptive genetic algorithms for solving job-shop scheduling problems[J]. Systems Engineering—Theory & Practice, 2004, 24(2): 58-62. (in Chinese)

[14] 张超勇, 饶运清, 李培根, 等. 柔性作业车间调度问题的两级遗传算法[J]. 机械工程学报, 2007, 43(4): 119-124.
ZHANG Chaoyong, RAO Yunqing, LI Peigen, et al. Bilevel genetic algorithm for the flexible job-shop scheduling problem[J]. Chinese Journal of Mechanical Engineering, 2007, 43(4): 119-124. (in Chinese)

改进 ADC 方法及其在武器装备系统效能评估中的应用*

刘仕雷¹, 李昊^{2,3}

(1. 装甲兵工程学院, 北京 100072; 2. 中国航天员科研训练中心 人因工程重点实验室, 北京 100094;
3. 清华大学 航天航空学院, 北京 100084)

摘要:在分析武器装备系统效能特点的基础上, 针对装备定型试验的应用需求, 提出一种改进可用性-可信性-能力法, 以解决传统方法在系统状态划分上无法准确全面描述武器装备实际可用性的问题。提出增加中间状态的策略, 对效能综合计算模式、状态描述与计算方法、状态转移计算方法和能力计算方法做整体改进, 在此基础上开发了支撑软件 SEEK。并以某型主战坦克系统效能评估为例, 验证了该方法的可行性和有效性。

关键词:武器装备; 定型试验; 系统效能; 评估; 主战坦克

中图分类号:E917 **文献标志码:**A **文章编号:**1001-2486(2017)03-130-06

Modified ADC method and its application for weapon system effectiveness evaluation

LIU Shilei¹, LI Hao^{2,3}

(1. Academy of Armored Force Engineering, Beijing 100072, China;
2. National Key Laboratory of Human Factors Engineering, China Astronaut Research and Training Center, Beijing 100094, China;
3. School of Aerospace Engineering, Tsinghua University, Beijing 100084, China)

Abstract: Based on analyses about system effectiveness of the weapon and aiming at application requirements from the weapon type approval tests, a modified ADC(availability-dependability-capability) method was put forward in order to solve the problem that the actual usability of the weapon cannot be presented accurately and entirely by dividing system conditions using the traditional ADC method. Comprehensive modifications were performed, such as putting forward the strategy of adding the medial conditions in the model. As the results showed, functions for integrating the sub-system effectiveness, for describing and computing the conditions, for transfers between conditions, and for capabilities were all improved. A software platform SEEK was developed based on these modifications. System effectiveness evaluation of a main tank was selected as an example for demonstrate feasibility and validity of the method.

Key words: weapon; type approval test; system effectiveness; evaluation; main tank

系统效能是随着系统结构和功能日益复杂、系统试验数据资源仍然相对缺乏和系统应用要求跨越式提高的综合作用的产物, 是建立在系统科学基础上的应用研究领域^[1]。

武器装备作为一类典型的复杂系统, 其系统效能的研究非常具有代表性^[2]。首先, 武器装备不论在结构还是功能上, 其复杂程度均呈现加速增长的趋势。先进的武器装备往往由多个功能各异又相互关联的分系统组成, 分系统又可划分为子系统, 子系统仍可继续划分, 从而形成结构上逐步分解、功能上相互关联的层次化体系结构。其次, 由于武器装备的应用环境是战场, 系统试验特

别是热试验往往难以实现, 从而导致在系统研发各生命阶段所需的数据资源都远远不能满足论证、研究和评估的需求。再次, 随着信息化和工业4.0时代的到来, 对战争的形态和内涵都将引发巨大变化, 由此带来对武器装备在应用要求方面的跨越式提高, 武器系统将不仅仅作为一种战争工具而独立存在。因此, 开展武器装备系统效能研究对理论方法和应用对象均具有重要意义^[3]。现阶段, 武器装备的定型试验是武器装备全生命周期的一个关键节点, 标志着武器装备由研发阶段进入应用阶段。因此本文选择这一阶段的武器装备作为研究对象。

* 收稿日期:2016-01-19
基金项目:国家自然科学基金资助项目(51375492,51575527);国家部委基金资助项目(SYFD130061815)
作者简介:刘仕雷(1976—),男,山东平度人,博士研究生,E-mail:zzblsl@163.com;
李昊(通信作者),男,副研究员,博士,E-mail:li-hao@tsinghua.edu.cn

目前,已有多种武器装备系统效能评估方法:按评估数据来源,可分为实装试验法、仿真模拟法、综合试验法、实战试验法等;按效能评估计算形式,可分为线性加权法、模糊综合评判法、层次分析法、概率综合法、灰色评估法等;按效能评估的模型,可以分为可用性 - 可信性 - 能力 (Availability-Dependability-Capability, ADC) 法、系统效能分析 (System Effectiveness Analysis, SEA) 法以及在传统方法基础上所做出的改进和扩展等。在实际应用中,往往需要根据评估的对象和目的,结合武器装备及其使用对象,在可行性和有效性等方面权衡,选择合适的评估方法^[4-5]。根据武器装备定型试验的应用需求,本文选择应用最为广泛的 ADC 法^[6] 作为基础,根据研究的需要对其各方面做了适应性修改,形成一种新的改进 ADC 法,并开发了软件工具 SEEK。

1 相关概念和方法

1.1 系统生命周期与效能评估

几乎所有人工系统都是根据需求反馈构想出的。系统全生命周期由四个串行且相互重叠的阶段构成,如图 1 所示,通过需求反馈将各阶段联成闭合回路^[7]。其中,测试过程是连接系统从研发到应用的重要环节。对武器装备来说,定型试验是整个测试过程的核心环节。

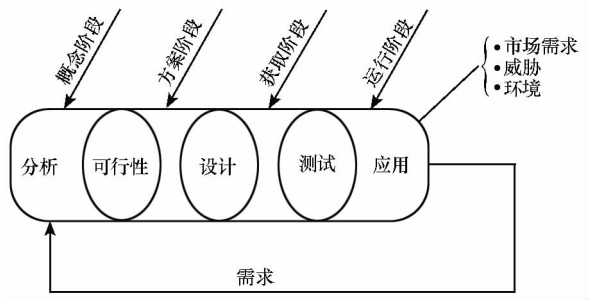


图 1 系统全生命周期的各阶段
Fig. 1 Phases of system life

图 2 给出了武器装备定型试验与研发、测评和应用三大回路之间的关系。

所谓系统的任务可靠性,是指系统凭借其功能,完成特定任务的能力。如主战坦克的火分系统,其任务可靠性是指在特定环境实现其任务(发射、将弹药到达目标并将其摧毁)的概率。系统的运行可靠性,则是指系统实现某种特定功能的能力,因此系统并非在所有运行时刻都打开。如主战坦克的信息分系统,只有在通信过程中才发挥作用。对于复杂的武器装备,如主战坦克,其

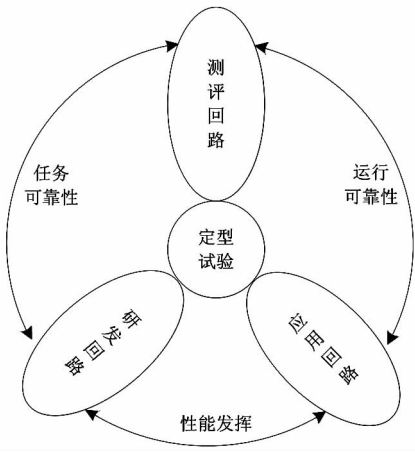


图 2 定型试验与三大回路的关系
Fig. 2 Relationship between type approval test & 3 main loops

中既涉及任务可靠性,又涉及运行可靠性。因此在研究中,将综合考虑这两类因素。

1.2 系统效能的定义及其特点分析

工程领域对系统效能的通用认识是:效能是一种预期的结果、输出、后果或操作,即正确地做正确的事情,以达到目的。系统效能是对系统完成其目标的能力的度量,是在系统预期能够完成的一组特定目标中的度量,也可被认为是系统能够实现其目标的概率。系统效能的要素包括:就绪状态(readiness)、设计完备度(design adequacy)和可靠性(reliability)^[1]。

在军事应用领域,被广泛认可的是美国武器系统效能咨询委员会 (Weapon System Effectiveness Industry Advisory Committee, WSEIAC) 于 1965 年提出的定义^[8]:系统效能是指系统能够完成一组给定任务的能力的度量,它是可用性、可信性和能力的函数。我国的 GJB 1394—1992 则将装备的效能定义为:在规定的条件下达到规定使用目标的能力。

上述两方面对系统效能要素的定义具有关联关系,如图 3 所示。

1.3 ADC 法的基本思想和应用局限

ADC 法是 WSEIAC 在其提出的系统效能定义的基础上建立的,其核心为系统效能模型。

$$E = A \cdot D \cdot C \tag{1}$$

其中, E 表示系统效能, A 表示系统可用性, D 表示系统可信性, C 表示系统能力。

ADC 法本质上是对武器装备系统主客观因素的综合,将系统效能与系统可用性、可信性和能力之间的关系用数学形式表示,在一定程度上简

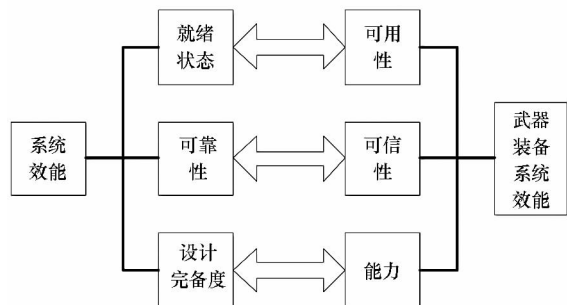


图 3 系统效能的要素

Fig. 3 Components of system effectiveness

化了评估的复杂程度;但在具体应用中,如何以简单的向量和矩阵形式完全描述系统各方面属性,对该方法提出了挑战。

传统的 ADC 法在本质上更适用于效能分析而非效能评估,也就是说,该方法更适用于武器装备的论证和设计阶段,而非装备的定型试验,原因如下。

1) 状态确定。在传统 ADC 法中,并没有对系统可用性向量 $A = [a_1 \ a_2 \ \cdots \ a_i \ \cdots \ a_n]$ 的中间状态给出确定方法,往往通过主观给出,或者进一步简化为省略中间状态的形式。另一方面,当将整个系统划分为多个层次后,系统状态数量将呈指数级增长,使得问题过于复杂而难以计算。

2) 状态转移。由于定型试验次数的限制,往往无法通过试验数据统计计算执行任务过程状态转移的概率,影响对系统可信性矩阵 D 的构建。此外,系统分层所导致的组合状态的增加,将增加计算复杂性。

3) 能力计算。对系统能力向量 C 的确定往往采用线性函数,这与装备实际情况并不符合;且当系统状态超过 2 个时,需要建立中间状态与能力之间的关系。

2 改进 ADC 法

2.1 综合计算模式的改进

传统 ADC 法对系统状态参数的确定,往往是在将武器装备看作一个大系统的假设下,主观给出的,随着系统复杂程度的增高,对状态的描述缺乏科学性。本文对综合计算模式的改进,在于解决系统效能的层次划分与聚合计算的问题。

2.1.1 系统效能层次划分

武器装备系统往往都是由很多结构和功能相对独立的成分构成,以这些独立成分单独作为研究对象时,其效能是可计算的。因此,在层次化和量子化思想的指导下,对武器装备做一定层次划

分(如图 4 所示),能够降低计算复杂度,并使得评估原始数据的准确性大大提升。



图 4 武器装备系统效能的层次划分

Fig. 4 Decomposition of weapon system effectiveness

层次划分须遵循以下原则:一是保持划分的成分在结构和功能上的完整性,能够独立开展效能计算;二是划分层次不宜过多。

2.1.2 基于塌陷效应的聚合计算方法

武器装备系统效能经过层次划分以后,每一层次上需要对所有成分的效能进行聚合计算,以获得上一层的系统效能。传统的聚合计算方法往往采用加权求和方法,该方法能够综合体现各成分效能对整体效能的影响,却无法描述系统效能的短板效应,即某一成分对系统整体效能具有很大影响的情况。为此,提出一种简单算法——基于塌陷效应的聚合计算方法,以解决短板效应问题。

基本思想:当聚合计算的分项值都“较高”(高于“临界值”)时,仍然按照传统的加权求和的方式进行聚合计算;当聚合计算的某个或某几个分项低于“临界值”时,需要考虑这种低值对整个系统效能或者系统能力的负面影响。

术语定义:所谓“塌陷效应”,其含义就是将各个分项效能看作是支撑整个系统效能的“支柱”,如果某个“支柱”的高度明显偏低,那么它将“拖累”或者“拖垮”与此相关的其他分项效能,同时也对整个系统的效能造成影响。

算法描述如下。

步骤 1: 加权求和。按传统加权求和方法计

算系统效能: $E = \sum_{i=1}^n (W_i \cdot E_i)$, 其中, n 为子系统数, E_i 为第 i 个分系统的效能, W_i 为相应权值。对权值的设置是历史信息、专家经验和实测结果的综合。例如,可使用文献[9]中提出的群体可拓层次分析法得到。

步骤 2: 确定临界值。根据试验结果与计算结果的比对,结合专家经验,为每个分系统效能设置临界值 K_i , 当分系统效能值低于该临界值时,对系统效能影响呈现剧烈变化。

步骤 3: 塌陷效应计算。若分系统效能满足 $E_i < K_i$, 则认为该分系统需进行塌陷效应计算。若有 m 个分系统需进行塌陷效应计算,则:

$$\begin{aligned}\bar{E} &= E \times \prod_{j=1}^m [1 - (K_j - E_j)] \\ &= \sum_{i=1}^n (W_i \cdot E_i) \times \prod_{j=1}^m [1 - (K_j - E_j)]\end{aligned}\quad (2)$$

得到的 \bar{E} 为系统效能值。

2.2 状态描述与计算方法的改进

为了建立系统故障模式与可识别状态之间的对应关系,引入故障模式群的概念。故障模式群由一定数量的故障模式组成,与可识别状态构成一一对应关系。如图5所示,由系统的 n 个故障模式计算 m 个可识别状态的概率值。显然,有: $Q_1 = \emptyset$ 且 $\sum_{i=2}^m q_i = n$ 。需要说明的是:对由多个分系统构成的系统来说,虽然故障模式与出故障的分系统之间往往是“多对多”的关系,为了研究方便,通常会做一些简化处理,使二者为简单的“一对多”关系。

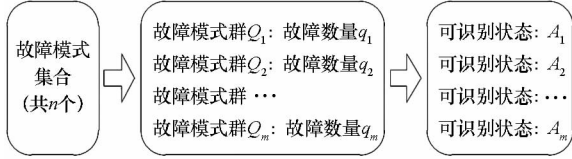


图5 故障模式到可识别状态的转化

Fig.5 Transition from failure patterns to identifiable states

令故障模式群 Q_i 中故障模式对应的平均故障间隔时间和平均故障修复时间,分别为 $MTBF_1, MTBF_2, \dots, MTBF_{q_i}$ 和 $MTTR_1, MTTR_2, \dots, MTTR_{q_i}$,

$$D = \begin{bmatrix} P(Q_2 + Q_3, t_i \geq T_1) & P(Q_2, t_i < T_1) \\ P(Q_2, t'_i < T_2) & 1 - d_{21} - d_{23} \\ P(Q_3, t'_i < T_2) & 1 - d_{31} - d_{33} \end{bmatrix} \begin{bmatrix} 1 - d_{11} - d_{12} \\ \frac{MTBF_{Q_3} - MTBF_{Q_2}}{MTBF_{Q_3}} \cdot P(Q_3, t_i < T_1) \\ P(Q_3, t'_i \geq T_2) \end{bmatrix}\quad (7)$$

其中: t_i 表示对每次故障时间的记录; t'_i 表示对每次维修时间的记录; $P(Q_2, t_i < T_1)$ 表示故障模式群 Q_2 中故障间隔时间小于规定时间 T_1 的比率,以此类推。

2.4 能力计算方法的改进

采用品质效用函数的方法进行能力指标的计算。武器装备常具备多个品质因素(性能),有的品质因素要求越大越好,有的要求越小越好,还有的要求在一定范围之内,而且不同的品质因素在装备中所发挥的作用也有差异。

对效用函数值做归一化处理。若武器装备有 m 个品质因素(性能),即 $d = (d_1, d_2, \dots, d_m)$, 性能指标最大值点 $d_{\max} = (d_{\max 1}, d_{\max 2}, \dots, d_{\max m})$, 最小值点 $d_{\min} = (d_{\min 1}, d_{\min 2}, \dots, d_{\min m})$ 。

若要求品质因素越大越好,则采用如式(8)

则故障模式群 Q_i 中所有故障模式的平均故障间隔时间与平均故障修复时间的加权平均值分别为:

$$\begin{cases} MTBF_{Q_i} = \frac{MTBF_1 + MTBF_2 + \dots + MTBF_{q_i}}{q_i} \\ MTTR_{Q_i} = \frac{MTTR_1 + MTTR_2 + \dots + MTTR_{q_i}}{q_i} \end{cases}\quad (3)$$

分以下三种情况计算系统状态概率。

1) 当系统处于故障状态时, $i = m$, 有:

$$A_m = \frac{MTTR_{Q_i}}{\sum MTBF_{Q_i} + \sum MTTR_{Q_i}}\quad (4)$$

2) 当系统处于中间状态时, $1 < i < m$, 有:

$$A_i = \frac{MTBF_{Q_i}}{\sum MTBF_{Q_i} + \sum MTTR_{Q_i}}\quad (5)$$

3) 当系统处于正常状态时, $i = 1$, 有:

$$A_1 = 1 - \sum_{i=2}^m A_i\quad (6)$$

2.3 状态转移计算方法的改进

在系统具有中间状态的情况下,系统状态共有 m 个,可信度矩阵(即状态转移矩阵)变为 m 阶矩阵。为了简化,仅考虑 $m = 3$ 的情况。

令 T_1 表示完成任务的时间(即状态转移参考量), T_2 表示允许维修的时间。根据对各故障模式的试验观测值,可以计算得到可信度矩阵中各元素的估值。

所示的效用函数。

$$\mu_k = d_k / d_{\max k}\quad (8)$$

若要求品质因素越小越好,则采用如式(9)所示的效用函数。

$$\mu_k = 1 + (d_{\min k} - d_k) / d_{\max k}\quad (9)$$

若要求品质因素在 $[r_1, r_2]$ 区间为宜,则采用如式(10)所示的效用函数。

$$\mu_k = \begin{cases} d_k / r_1 & d_k \in [d_{\min k}, r_1] \\ 1 & d_k \in [r_1, r_2] \\ 1 + (r_2 - d_k) / d_{\max k} & d_k \in [r_2, d_{\max k}] \end{cases}\quad (10)$$

得到品质因素效用函数值的计算结果为 $\mu = (\mu_1, \mu_2, \dots, \mu_m)$ 。通过对 μ 中各元素的加权求和,结合基于塌陷效应的聚合计算方法,可以得到正常状态的能力值 c_1 。显然有故障状态能力值

$c_m = 0$ 。通过专家经验,得到降额系数 ϕ_i ,以计算中间状态的能力值 $c_i = \phi_i \cdot c_1$ 。

2.5 支撑软件工具 SEEK

在前述四个方面算法改进基础上,应用通用程序开发平台研制了武器装备系统效能评估支撑软件工具 SEEK,其主要功能如图 6 所示,主要包括四个部分:基本功能模块提供了软件运行和使用的基本功能和界面,并预留了其他功能模块(如仿真模块)的接口;模板模块提供了研究对象,即各类武器装备的层次化结构模板,并允许用户根据需要自定义新模板和修改已有模板;算法模块是软件的核心部分,分别实现了改进 ADC 法的所有功能;数据库模块是软件的基础,包含了定型试验数据、武器装备系统各组成部分的设计数据以及相关的专家经验和历史数据。

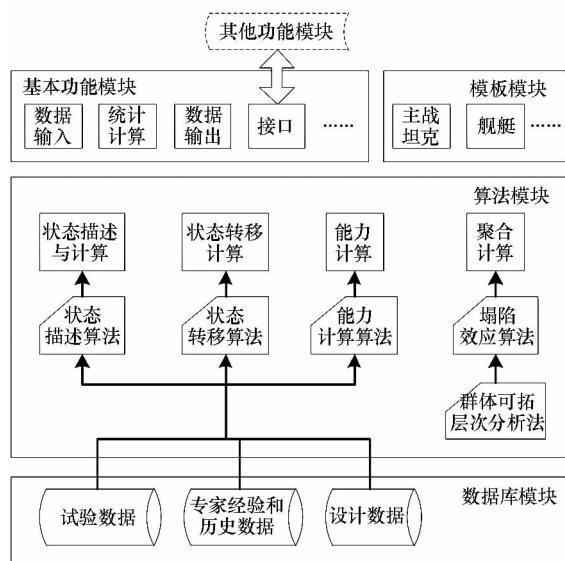


图 6 SEEK 的主要功能组成

Fig. 6 Main functional components of SEEK

3 主战坦克系统效能评估

为了演示验证改进 ADC 法的应用情况,本节选择某型主战坦克作为系统效能评估研究的研究案例^[10-11]。

3.1 系统描述与基本假设

SEEK 软件从效能评估的角度,将主战坦克看作由火力分系统、机动分系统、防护分系统、信息分系统、电气分系统构成的综合系统,并考虑人因工程影响。因此,主战坦克的整体效能为:

$$E_T = (W_f E_f + W_m E_m + W_p E_p + W_e E_e + W_c E_c) C_h \quad (11)$$

其中, E_f 、 E_m 、 E_p 、 E_e 和 E_c 分别是火力分系统、机动分系统、防护分系统、信息分系统、电气分系统

的效能, W_f 、 W_m 、 W_p 、 W_e 和 W_c 分别是各分系统对应的权值, C_h 表示人因工程影响。

主战坦克火力分系统继续分解,又可分为火控子系统、火炮子系统、自动装弹子系统和辅助武器子系统。其中,火控子系统由火控计算机、各种传感器、瞄准镜、激光测距仪、微夜视装置、火炮稳定装置、火炮控制装置组成。

主战坦克系统效能评估是结合其设计定型基地试验进行的,其数据来源包括定型试验数据和经验数据,这些数据都被存入 SEEK 中。

3.2 火控子系统效能评估

下面以主战坦克的火控子系统为代表,介绍系统效能评估的分析过程。

根据改进 ADC 法,火控子系统效能计算模型为:

$$E_{fa} = A_{fa} \cdot D_{fa} \cdot C_{fa} \quad (12)$$

根据改进 ADC 法的步骤,开展系统效能评估计算。

火控子系统可用性矩阵 $A_{fa} = [A_{fa1} \ A_{fa2} \ A_{fa3}]$ 。为了简化,仅考虑三种状态,即:

- 1) 完好状态概率 A_{fa1} ,描述无故障模式;
- 2) 中间状态概率 A_{fa2} ,描述轻微故障模式,与火控计算机、各种传感器、瞄准镜、激光测距仪四类装置的失效相关,构成故障模式群 Q_2 ;
- 3) 故障状态概率 A_{fa3} ,描述故障状态模式,与微夜视装置、火炮稳定装置、火炮控制装置三类装置的失效相关,构成故障模式群 Q_3 。

根据火控子系统七类装置的故障时间和故障维修时间的试验结果,使用式(3)~(6),计算得到:

$$A_{fa} = [A_{fa1} \ A_{fa2} \ A_{fa3}] = [0.656 \ 0.263 \ 0.081] \quad (13)$$

在对火控子系统可信度矩阵计算中,令任务时间 $T_1 = 4$ h,允许维修时间 $T_2 = 1$ h。根据试验观测结果,使用式(7),在 SEEK 支持下计算得到:

$$D_{fa} = \begin{bmatrix} 0.78 & 0.12 & 0.10 \\ 0.90 & 0.04 & 0.06 \\ 0.84 & 0.00 & 0.16 \end{bmatrix} \quad (14)$$

使用式(8)~(10),利用专家经验,在 SEEK 支持下计算得到:

$$C_{fa} = [0.731 \ 9 \ 0.292 \ 8 \ 0]^T \quad (15)$$

将结果代入式(12)中,得到 $E_{fa} = 0.626 \ 3$ 。

3.3 主战坦克系统效能计算

表 1 给出火力分系统中各子系统的效能值和权值。其中:各子系统效能值是根据 3.2 节的过程分别计算得到的;权值是使用 SEEK 中群体可

拓层次分析法模块,根据历史信息、专家经验和实测结果的综合计算得出的。此外,根据专家经验,得到每个子系统的临界值 K_i 均为 0.6,因此对自动装弹子系统和辅助武器子系统需要启动基于塌陷效应的聚合计算,根据式(2),使用 SEEK 计算得到火力分系统的效能值 $E_i=0.479\ 1$ 。

表 1 火力分系统中各子系统的效能计算数据

Tab.1 Data for calculating MOE of elements of firepower subsystem

	子系统			
	火控子系统	火炮子系统	自动装弹子系统	辅助武器子系统
效能值	0.626 3	0.738 1	0.525 6	0.508 3
权值	0.25	0.12	0.26	0.37

同理,针对其他分系统,根据火力分系统的计算步骤计算其效能值,并仍使用群体可拓层次分析法计算相应权值,结果如表 2 所列。另有 $C_h=0.943\ 7$,使用式(11),在 SEEK 中最终得到主战坦克系统效能值 $E_T=0.569\ 0$ 。

表 2 主战坦克各分系统的效能计算数据

Tab.2 Data for calculating MOE of subsystems of main tank

	分系统				
	火力分系统	机动分系统	防护分系统	信息分系统	电气分系统
效能值	0.479 1	0.678 2	0.737 2	0.639 2	0.508 2
权值	0.257 3	0.269 6	0.139 4	0.186 7	0.147 0

为进一步验证方法及其应用的可信性,利用 SEEK 平台,将研究结果与研究对象的热试验数据进行对比,可以发现二者间具有相似规律。从而也进一步验证了该方法的有效性。

4 结论

改进 ADC 方法通过在武器装备的正常状态和故障状态之间增加中间状态的策略,提高了对系统状态描述的准确性;同时,通过对效能综合计算模式、状态描述与计算方法、状态转移计算方法和能力计算方法的整体改进,一定程度上解决了传统方法过于依赖专家主观评价的困难,有效提升了效能评估应用的适应性和准确性。该方法在对某型主战坦克系统效能评估应用中得到了验证。

改进 ADC 方法尚未全面考虑武器装备使用

者对武器装备系统效能发挥的作用和影响,也未涉及实际作战想定,特别是对抗情况下武器装备系统效能的评估,这两方面的问题有待在后续研究中解决。

参考文献(References)

[1] Habayeb A R. System effectiveness [M]. UK: Pergamon Press, 1987.

[2] 杨峰, 王维平. 武器装备作战效能仿真与评估[M]. 北京: 电子工业出版社, 2014.
YANG Feng, WANG Weiping. Weapons and equipment operational effectiveness simulation and evaluation [M]. Beijing: Publishing House of Electronics Industry, 2014. (in Chinese)

[3] Li Z H, Zhang S, Wang J Y, et al. Research on description method of operational task oriented to operational effectiveness evaluation [C] //Proceedings of AsiaSim, Communications in Computer and Information Science, 2012, 3: 313 - 321.

[4] 梁金登, 李东旭. 空间武器系统效能分析研究[J]. 火力与指挥控制, 2009, 34(5): 47 - 50.
LIANG Jindeng, LI Dongxu. Study on effectiveness evaluation method of space weapon system[J]. Fire Control & Command Control, 2009, 34(5): 47 - 50. (in Chinese)

[5] 郭齐胜, 张磊. 武器装备系统效能评估方法研究综述[J]. 计算机仿真, 2013, 30(8): 1 - 4.
GUO Qisheng, ZHANG Lei. Research summary of weapons equipment systems effectiveness evaluation methods [J]. Computer Simulation, 2013, 30(8): 1 - 4. (in Chinese)

[6] 侯立峰, 熊哲, 盛景军. 基于 ADC 方法的飞行保障系统效能评估模型[J]. 火力与指挥控制, 2010, 35(10): 123 - 126.
HOU Lifeng, XIONG Zhe, SHENG Jingjun. Model of efficiency evaluation on flight support system based on ADC method [J]. Fire Control & Command Control, 2010, 35(10): 123 - 126. (in Chinese)

[7] Zacks S. Introduction to reliability analysis: probability models and statistical methods [M]. US: Springer-Verlag New York Inc., 1992.

[8] Weapon System Effectiveness Industry Advisory Committee. Final report of task group 2: prediction measurement; AFSC TR - 65 - 2[R]. US: Weapon System Effectiveness Industry Advisory Committee, 1965: 1 - 3.

[9] 孟庆均, 宋爱斌, 朱立民. 群体可拓层次分析法在 C⁴ISR 系统效能指标赋权中的应用[J]. 装甲兵工程学院学报, 2008, 22(2): 25 - 29.
MENG Qingjun, SONG Aibin, ZHU Limin. GEHP and application in determining the effectiveness targets weight of C⁴ISR system [J]. Journal of Academy of Armored Force Engineering, 2008, 22(2): 25 - 29. (in Chinese)

[10] 罗来科, 蒋宝唐, 宣益民. 主战坦克作战效能分析与模糊综合评价[J]. 火力与指挥控制, 2003, 28(6): 39 - 41.
LUO Laike, JIANG Baotang, XUAN Yimin. The efficiency analysis of the main battle tank and its fuzzy synthetic evaluation [J]. Fire Control & Command Control, 2003, 28(6): 39 - 41. (in Chinese)

[11] 王正元, 刘靖旭, 谭跃进, 等. 基于仿真的主战坦克作战效能评估方法[J]. 计算机仿真, 2005, 22(1): 29 - 32.
WANG Zhengyuan, LIU Jingxu, TAN Yuejin, et al. Combat effectiveness evaluation to armor based on combat simulation[J]. Computer Simulation, 2005, 22(1): 29 - 32. (in Chinese)

干扰条件下基于空频域二次优化的 MIMO 雷达波形设计方法*

王玉玺¹, 黄国策¹, 李 伟¹, 胡继宽²

(1. 空军工程大学 信息与导航学院, 陕西 西安 710077; 2. 空军大连通信士官学校, 辽宁 大连 116600)

摘 要:针对干扰条件下多输入多输出雷达发射方向图优化问题,提出一种基于空频域二次优化的多输入多输出雷达波形设计方法。该方法将空域上方向图优化问题转化为关于空时发射序列协方差矩阵的优化问题,利用多输入多输出雷达发射方向图仅与阵元之间波形相关性有关的特性,进一步降低空域波形设计复杂度,并通过 p 阶导数约束展宽零陷;针对优化得到的协方差矩阵,利用随机向量法通过最小二乘准则逼近最优发射方向图来合成具体恒包络波形;在基于空域优化得到的发射波形基础上,根据改变不同时刻信号序列的初始相位雷达发射方向图不变的特性,通过拟功率方法优化相位变化矩阵,实现雷达波形在频域上的二次优化以抑制频域上的干扰。仿真实验证明了所提方法在方向图匹配和干扰抑制方面的有效性。

关键词:多输入多输出雷达;波形设计;干扰抑制;恒包络
中图分类号:TN911.7 **文献标志码:**A **文章编号:**1001-2486(2017)03-136-08

MIMO radar waveform design method based on quadratically spatial and spectral optimizations under jamming

WANG Yuxi¹, HUANG Guoce¹, LI Wei¹, HU Jikuan²

(1. Information and Navigation College, Air Force Engineering University, Xi'an 710077, China;
2. Dalian Air Force Communication Noncommissioned Officer Academy, Dalian 116600, China)

Abstract: For the optimization problem of MIMO (multiple input multiple output) radar transmit beampattern under jamming, a new MIMO radar waveform design method based on quadratically spatial and spectral optimization was proposed. Firstly, the proposed method converted the problem of MIMO radar transmit beampattern design into the optimization problem about the covariance matrix of MIMO radar's transmit space-time sequences. Based on the fact that MIMO radar transmit beampattern was only decided by the correlation of each element's transmit waveforms, the computational burden of the spatial optimization of waveforms could be reduced. Furthermore, the nulling towards the jamming direction of the transmit beampattern was broadened by the p -order derivative constraint. With the optimized covariance matrix, the randomization method was used to synthesize the actual constant modular waveforms under the criteria of least square to gain on the optimal beampattern. Finally, with the optimized waveforms through spatial optimizing process, a phase flexible diagonal matrix was optimized with the like-power method to achieve the spectral optimization of MIMO radar waveforms based on the fact that MIMO radar transmit beampattern would not be influenced by the change of the initial phase of transmit sequence at a certain moment. And the spectral jamming could be avoided by the spectral optimization of waveforms. Simulation results prove the effectiveness of the proposed method in matching desired beampattern and anti-jamming.

Key words: multiple input multiple output radar; waveform design; anti-jamming; constant envelop

多输入多输出 (Multiple Input Multiple Output, MIMO) 雷达凭借每个阵元能够发射不同波形的优异性能受到广泛关注^[1-9]。根据 MIMO 雷达阵元布置以及信号处理的特点,可将其分为分布式 MIMO 雷达和集中式 MIMO 雷达。其中分布式 MIMO 雷达通过空间分集可以有效消除目标闪烁带来的影响^[1-2];而集中式 MIMO 雷达则利用波形分集形成较大的虚拟阵列孔径,提高雷达参数估计、目标识别和干扰抑制等性能^[3],本文主要研究集中式 MIMO 雷达。

传统集中式 MIMO 雷达,每个阵元通过发射相互正交信号,发射端发射功率在空间均匀分布^[3]。为提高雷达发射功率的利用率,利用不同阵元之间发射波形的相关性设计 MIMO 雷达发射方向图,实现发射功率在特定空域范围内的聚焦已成为目前研究的热点^[4-10]。现有关于 MIMO

* 收稿日期:2016-09-18
基金项目:国家自然科学基金资助项目(61302153)
作者简介:王玉玺(1989—),男,山东寿光人,博士研究生,E-mail:WYX10013@163.com;
黄国策(通信作者),男,教授,博士,博士生导师,E-mail:huangguoce@163.com

雷达发射方向图设计的波形优化方法可分为两步,首先根据期望发射方向图优化发射波形协方差矩阵,然后利用优化协方差矩阵匹配设计具体发射波形。文献[4]首次推导了 MIMO 雷达发射方向图计算公式,建立了发射方向图优化模型并利用梯度算法求解发射波形的协方差矩阵;文献[5]则提出了经典的方向图匹配设计模型和最小化旁瓣方向图设计模型;文献[6]在文献[4]的基础上提出了一种关于协方差矩阵的无约束半正定规划模型;文献[7-8]为避免直接优化协方差矩阵,分别提出了无约束实相关矩阵综合方法;为降低协方差矩阵优化计算复杂度,文献[9-10]分别提出了基于发射加权矩阵优化的 MIMO 雷达发射方向图优化算法,将 MIMO 雷达波形设计问题转化为关于正交基波形加权矩阵的优化问题;文献[11]则在文献[9]基础上进一步研究了优化波形的模糊函数。文献[12]针对主瓣波动和旁瓣电平进行了研究;文献[13]在现有方向图匹配准则的基础上进一步推广,提出一种旁瓣控制方向图设计方法;通过上述不同方法对发射波形协方差矩阵进行优化后,接下来则需要根据优化后的协方差矩阵设计具体的发射波形。由于在实际应用中天线阵元发射功率放大器具有非线性特性,因此为保证发射波形不失真并最大化功率利用率,需要发射波形满足恒包络特性。目前最为通用的波形设计方法为文献[14]所提基于协方差矩阵匹配的循环算法,该方法虽然能够以闭合解的形式给出具体的发射波形,但是算法为高度非凸非线性优化问题对初始迭代点非常敏感,而且该方法在波形合成时没有考虑干扰情况下雷达发射方向图的置零约束,因此优化后的波形不能保证雷达发射方向图在干扰方向上形成满足要求的零陷。

现有文献大都针对理想情况下 MIMO 雷达发射方向图及波形优化设计进行研究,而没有考虑实际应用中特别是在复杂电磁环境下, MIMO 雷达不仅可能面临来自空域特定方向的干扰,而且还有可能在频域上面临来自敌方甚至是己方与雷达具有重叠频带的其他无线电设备的干扰。本文针对上述问题,提出一种干扰条件下基于空频域二次优化的 MIMO 雷达波形设计方法。

1 MIMO 雷达信号模型

设集中式 MIMO 雷达发射阵列为一均匀线

阵,阵元数目为 M 且阵元间距为 $d = \frac{f_0}{2c}$, f_0 为发射信号载频, c 为光速。设在 n 时刻 M 个阵元发射基带离散信号序列为:

$$\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_M(n)]^T \in C^M \quad (1)$$

则 MIMO 雷达在一次相干处理间隔内发射基带离散信号矩阵可表示为:

$$\mathbf{S} = [s(0), s(1), \dots, s(N-1)] = [s_1, s_2, \dots, s_M]^T \in C^{M \times N} \quad (2)$$

其中, $\mathbf{s}_m = [s_m(0), s_m(1), \dots, s_m(N-1)]^T \in C^N$ 表示第 m 个阵元发射的信号序列, N 为一次相干处理时间内信号取样次数即信号码长,由信号带宽和发射脉冲宽度决定。假设各个阵元发射的波形均为窄带信号,则在 n 时刻,远场 θ 方向接收到的信号为:

$$r(n, \theta) = \mathbf{a}^H(\theta) \mathbf{s}(n) \quad (3)$$

其中, $\mathbf{a}(\theta) = [1, e^{-j\pi \sin(\theta)}, \dots, e^{-j(M-1)\pi \sin(\theta)}]^T$ 为发射阵列导向矢量。因此在 n 时刻 MIMO 雷达发射波形在空间的能量分布为:

$$P(n, \theta) = \mathbf{a}^H(\theta) E\{s(n)s^H(n)\} \mathbf{a}(\theta) = \mathbf{a}^H(\theta) \mathbf{R} \mathbf{a}(\theta) \quad (4)$$

其中, $\mathbf{R} = E\{s(n)s^H(n)\}$ 表示 n 时刻雷达发射信号的协方差矩阵。由于 \mathbf{R} 为雷达发射信号协方差矩阵的理论值,在统计理论上满足如下关系式:

$$\mathbf{R} = \frac{1}{N} \sum_{n=1}^N s(n)s^H(n) = \frac{1}{N} \mathbf{S} \mathbf{S}^H = \frac{1}{N} \mathbf{R} \quad (5)$$

其中 $\mathbf{R} = \mathbf{S} \mathbf{S}^H$ 。在相干处理时间内, MIMO 雷达发射方向图可表示为:

$$\begin{aligned} P(\theta) &= \sum_{n=0}^{N-1} \mathbf{a}^H(\theta) E\{s(n)s^H(n)\} \mathbf{a}(\theta) \\ &= \mathbf{a}^H(\theta) \mathbf{R} \mathbf{a}(\theta) = \mathbf{a}^H(\theta) \mathbf{S} \mathbf{S}^H \mathbf{a}(\theta) \end{aligned} \quad (6)$$

将整个空域 $\Theta = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ 划分为 L 个离散

点,雷达期望发射方向图为 $P_d(\theta_l)$, $\theta_l \in \Theta$ 。假设在空域 θ_c 方向存在一干扰,则为了抑制该干扰需要使雷达发射方向图在 θ_c 方向上形成零陷,即:

$$\mathbf{a}^H(\theta_c) \mathbf{s}(n) = 0, n = 1, 2, \dots, N \quad (7)$$

利用方向图匹配准则^[5]可以得到带有零陷约束的协方差矩阵优化模型为:

$$\begin{cases} \min_{\mathbf{R}, r} \sum_{l=1}^L |r P_d(\theta_l) - \mathbf{a}^H(\theta_l) \mathbf{R} \mathbf{a}(\theta_l)|^2 \\ \text{s. t.} \quad \mathbf{R}(m, m) = \frac{E}{M}, m = 1, 2, \dots, M \\ \mathbf{a}^H(\theta_c) \mathbf{R} \mathbf{a}(\theta_c) = 0 \\ \mathbf{R} \geq 0 \end{cases} \quad (8)$$

其中, r 为尺度因子, E 表示雷达总的发射功率。

式(8)为关于 \mathbf{R} 的半正定规划问题,可以利用 CVX 工具箱高效求解。在求得最优协方差矩阵 \mathbf{R} 后,需要根据 $\mathbf{R} = \mathbf{S}\mathbf{S}^H$ 合成具体的发射波形,利用文献[14]所提循环优化方法可以得到具体的发射波形,即:

$$\begin{cases} \min_{\mathbf{S}} \|\mathbf{S} - \mathbf{R}^{1/2}\mathbf{U}\| \\ \text{s. t. } |\mathbf{S}(m, n)|^2 = \frac{E}{MN} \quad m = 1, \dots, M \quad n = 1, \dots, N \end{cases} \quad (9)$$

其中, $\mathbf{R}^{1/2}$ 表示协方差矩阵的 Hermite 均方根, $\mathbf{U} \in \mathbb{C}^{M \times N}$ 为半正交矩阵,约束条件表示雷达发射波形为具有恒包络特性。利用式(8)和式(9)虽然能够解决 MIMO 雷达发射波形设计问题,但是仍然存在几点不足:①在考虑发射波形恒包络或低峰均值比(Peak-to-Average-power Ratio, PAR)等实际约束条件下,基于协方差矩阵匹配的波形设计是一个高度非凸非线性优化问题,而且循环算法对初始迭代点非常敏感;②由于循环算法在合成雷达波形矩阵 \mathbf{S} 时,仅以最小二乘准则逼近矩阵 $\mathbf{R}^{1/2}\mathbf{U}$,而没有考虑零陷约束,因此不能保证优化波形在方向 θ_c 处形成满足条件的零陷;③在实际应用中特别是复杂电磁环境下, MIMO 雷达不仅面临来自空域的具有特定方向的杂波干扰,而且还可能面临来自敌方特定频谱上的干扰,甚至是己方与雷达工作频段相重叠的其他无线电设备频域上的干扰,而现有关于 MIMO 雷达波形设计方法仅从空域对雷达波形进行优化,无法同时抑制来自空域和频域上的干扰。

2 基于空频域二次优化的 MIMO 雷达波形设计

针对复杂电磁环境下 MIMO 雷达有可能同时面临来自空域和频域干扰的情况,设计一种基于空频域二次优化的 MIMO 雷达波形设计方法,通过分别在空域和频域内对雷达波形进行优化,在匹配期望发射方向图的条件下,同时抑制来自空域和频域的干扰。

2.1 基于空域的 MIMO 雷达波形设计

虽然式(7)可以保证 MIMO 雷达最优发射方向图在干扰 θ_c 方向上形成一零陷,但是所得零陷较窄,无法保证雷达与干扰源相对移动时干扰始终处于零陷内,为提高雷达干扰抑制的有效性,可利用 p 阶导数约束方法对干扰零陷展宽,即:

$$\left. \frac{\partial^r (\mathbf{a}^H(\theta) \mathbf{s}(n))}{\partial \xi^r} \right|_{\theta=\theta_c} = c_r (\mathbf{D}' \mathbf{a}(\theta_c))^H \mathbf{s}(n) = 0 \quad (10)$$

其中: $n = 1, \dots, N; r = 1, \dots, p; \xi = \pi \sin(\theta); c_r = j^r (\sum_{m=1}^M z_m^{2r})^{1/2}; \mathbf{D}' = (\sum_{m=1}^M z_m^{2r})^{-1/2} \text{diag}([z_1^r, z_2^r, \dots, z_M^r]); z_m = m - 1, m = 1, \dots, M$ 。令 $\mathbf{x} = \text{vec}(\mathbf{S})$, $\mathbf{X} = \mathbf{x}\mathbf{x}^H$, 由式(6)可知:

$$\begin{aligned} P(\theta) &= \mathbf{a}^H(\theta) \mathbf{S} \mathbf{S}^H \mathbf{a}(\theta) \\ &= (\mathbf{I}_N \otimes \mathbf{a}^H(\theta) \text{vec}(\mathbf{S}))^H (\mathbf{I}_N \otimes \mathbf{a}^H(\theta) \text{vec}(\mathbf{S})) \\ &= \text{vec}(\mathbf{S})^H \mathbf{I}_N \otimes \mathbf{a}(\theta) \mathbf{I}_N \otimes \mathbf{a}^H(\theta) \text{vec}(\mathbf{S}) \\ &= \text{tr}(\mathbf{I}_N \otimes \mathbf{a}^H(\theta) \text{vec}(\mathbf{S}) \text{vec}(\mathbf{S})^H \mathbf{I}_N \otimes \mathbf{a}(\theta)) \\ &= \text{tr}(\mathbf{I}_N \otimes \mathbf{a}^H(\theta) \mathbf{X} \mathbf{I}_N \otimes \mathbf{a}(\theta)) \\ &= \text{tr}(\mathbf{A}^H(\theta) \mathbf{X} \mathbf{A}(\theta)) \\ &= \text{tr}(\mathbf{V}(\theta) \mathbf{X}) \end{aligned} \quad (11)$$

其中,“ \otimes ”表示 Kronecker 乘积运算, $\mathbf{A}(\theta) = \mathbf{I}_N \otimes \mathbf{a}(\theta)$, $\mathbf{V}(\theta) = \mathbf{A}(\theta) \mathbf{A}^H(\theta)$ 。式(11)推导中利用了矩阵向量化和 Kronecker 乘积运算特性以及矩阵迹运算特性。同理,零陷展宽约束式(10)可表示为:

$$\begin{aligned} &(\mathbf{D}' \mathbf{a}(\theta_c))^H \mathbf{S} \mathbf{S}^H \mathbf{D}' \mathbf{a}(\theta_c) \\ &= \text{vec}((\mathbf{D}' \mathbf{a}(\theta_c))^H \mathbf{S})^H \text{vec}((\mathbf{D}' \mathbf{a}(\theta_c))^H \mathbf{S}) \\ &= \text{vec}(\mathbf{S})^H \mathbf{I}_N \otimes ((\mathbf{D}' \mathbf{a}(\theta_c)) (\mathbf{D}' \mathbf{a}(\theta_c))^H) \text{vec}(\mathbf{S}) \\ &= \mathbf{x}^H \mathbf{H}(\theta_c) \mathbf{x} = \text{tr}(\mathbf{H}(\theta_c) \mathbf{X}) \end{aligned} \quad (12)$$

其中 $\mathbf{H}(\theta_c) = \mathbf{I}_N \otimes ((\mathbf{D}' \mathbf{a}(\theta_c)) (\mathbf{D}' \mathbf{a}(\theta_c))^H)$ 。因此带有展宽零陷的 MIMO 雷达发射方向图设计问题可转化为关于协方差矩阵 \mathbf{X} 的优化问题 P_1 , 即:

$$\begin{cases} \min_{\alpha, \mathbf{X}} \sum_{l=1}^L |\alpha P_d(\theta_l) - \text{tr}(\mathbf{V}(\theta_l) \mathbf{X})|^2 \\ \text{s. t. } \text{tr}(\mathbf{H}(\theta_c) \mathbf{X}) \leq \varepsilon \\ \text{diag}(\mathbf{X}) = \frac{E}{MN} \\ \text{rank}(\mathbf{X}) = 1, \mathbf{X} \geq 0 \end{cases} \quad (13)$$

其中, ε 表示零陷深度, α 表示尺度因子用于更好地匹配期望方向图,第二个约束条件表示每个阵元发射波形为恒包络的。优化问题 P_1 可以通过半正定松弛忽略阶为 1 的约束条件,将非凸问题转化为凸的半正定规划问题并求得最优协方差矩阵 \mathbf{X} 。但是直接通过 P_1 求解协方差矩阵 \mathbf{X} 计算复杂度为 $O((MN)^{3.5})$,特别是当发射波形码元数目 N 较大时,不能满足雷达发射波形优化的实时性要求。为降低计算复杂度,根据式(6)可知:

$$\begin{aligned} &\mathbf{a}^H(\theta) \mathbf{S} \mathbf{S}^H \mathbf{a}(\theta) \\ &= \mathbf{N} \mathbf{a}^H(\theta) \bar{\mathbf{R}} \mathbf{a}(\theta) \\ &= \text{tr}(\mathbf{I}_N \otimes \mathbf{a}^H(\theta) \mathbf{I}_N \otimes \bar{\mathbf{R}} \mathbf{I}_N \otimes \mathbf{a}(\theta)) \\ &= \text{tr}(\mathbf{A}^H(\theta) \mathbf{I}_N \otimes \bar{\mathbf{R}} \mathbf{A}(\theta)) \end{aligned} \quad (14)$$

由式(11)和式(14)对比可知,在空域上通过

优化 MIMO 雷达空时序列协方差矩阵 \mathbf{X} 设计 MIMO 雷达发射方向图等效于对矩阵 $\mathbf{I}_N \otimes \bar{\mathbf{R}}$ 的优化。由于 MIMO 雷达发射方向图仅由各阵元发射波形之间的相关性决定,而与码元序列之间的相位差无关,在空域上优化雷达空时序列协方差矩阵 \mathbf{X} 等价于对 n 时刻雷达发射信号的协方差矩阵 $\bar{\mathbf{R}} = E\{s(n)s^H(n)\}$ 的优化。因此基于空域的 MIMO 雷达发射方向图优化模型 P_2 可表示为:

$$\begin{cases} \min_{\mathbf{R}, r} \sum_{l=1}^L |rP_d(\theta_l) - \mathbf{a}^H(\theta_l)\bar{\mathbf{R}}\mathbf{a}(\theta_l)|^2 \\ \text{s. t.} \quad \text{diag}(\bar{\mathbf{R}}) \leq \gamma \frac{E}{MN} \\ (D^r\mathbf{a}(\theta_c))^H \bar{\mathbf{R}} D^r\mathbf{a}(\theta_c) \leq \varepsilon \\ \text{tr}(\bar{\mathbf{R}}) = \frac{E}{N} \\ \text{rank}(\bar{\mathbf{R}}) = 1 \\ \bar{\mathbf{R}} \geq \mathbf{0} \end{cases} \quad (15)$$

需要注意的是问题 P_2 为对 n 时刻发射波形协方差矩阵 $\bar{\mathbf{R}} = E\{s(n)s^H(n)\}$ 的优化,而非式(8)中的 $\mathbf{R} = \mathbf{S}\mathbf{S}^H$,因此在 P_2 中波形功率约束为 $\text{tr}(\bar{\mathbf{R}}) = \frac{E}{N}$ 且 $\text{rank}(\bar{\mathbf{R}}) = 1$ 。此外, P_2 中第一个约束条件表示对 n 时刻每个阵元的发射功率进行约束,其中 $\gamma \in [1, M]$ 。由于矩阵 $\bar{\mathbf{R}}$ 的阶 1 约束条件, P_2 为非凸的,因此可利用半正定松弛方法省略矩阵 $\bar{\mathbf{R}}$ 的阶 1 约束条件,将 P_2 松弛变换为 P_3 ,即:

$$\begin{cases} \min_{\mathbf{R}, r} \sum_{l=1}^L |rP_d(\theta_l) - \mathbf{a}^H(\theta_l)\bar{\mathbf{R}}\mathbf{a}(\theta_l)|^2 \\ \text{s. t.} \quad \text{diag}(\bar{\mathbf{R}}) \leq \gamma \frac{E}{MN} \\ (D^r\mathbf{a}(\theta_c))^H \bar{\mathbf{R}} D^r\mathbf{a}(\theta_c) \leq \rho \\ \text{tr}(\bar{\mathbf{R}}) = \frac{E}{N}, \bar{\mathbf{R}} \geq \mathbf{0} \end{cases} \quad (16)$$

其中 P_3 为关于 $\bar{\mathbf{R}}$ 的半正定规划问题,可以通过 CVX 工具箱高效求解。在求得优化矩阵 $\bar{\mathbf{R}}^*$ 后,相应的 $\mathbf{I}_N \otimes \bar{\mathbf{R}}^*$ 可直接作为 MIMO 雷达空时序列的最优协方差矩阵 \mathbf{X}^* 。由于 $\mathbf{X}^* = \mathbf{I}_N \otimes \bar{\mathbf{R}}^*$,因此在优化 $\bar{\mathbf{R}}$ 时,若 $\gamma = 1$,则 MIMO 雷达各个阵元发射功率相同而且每个阵元发射波形满足恒包络;若 $1 < \gamma \leq M$,则 MIMO 雷达不同阵元之间发射功率具有一定变化范围,但是每个阵元发射波形仍然满足恒包络。

通过优化问题 P_3 求得 $\bar{\mathbf{R}}^*$ 并根据 $\mathbf{I}_N \otimes \bar{\mathbf{R}}^*$ 获得最优协方差矩阵 \mathbf{X}^* 后,接下来则需要根据 \mathbf{X}^* 设计具体的发射波形。若优化后的协方差矩阵

\mathbf{X}^* 满足 $\text{rank}(\mathbf{X}^*) = 1$,经过特征值分解后,其非零特征值所对应的特征向量即为期望的恒包络发射波形;但是由于优化问题 P_3 经过松弛变化,实际所得优化矩阵 \mathbf{X}^* 的阶往往大于 1,此时可以利用随机向量合成方法^[15]得到满足约束条件的恒包络发射波形,其具体求解过程为:当 $\text{rank}(\mathbf{X}^*) \geq 2$ 时,任意选取 Q 个随机向量 \mathbf{x}_q ,且 \mathbf{x}_q 服从均值为 $\mathbf{0}$ 方差为 \mathbf{X}^* 的复高斯正态分布,即 $\mathbf{x}_q \sim N_c(\mathbf{0}, \mathbf{X}^*)$, $q = 1, 2, \dots, Q$,其中 Q 为随机化实验次数。

计算 $\mathbf{y}_q = \sqrt{\frac{E}{MN}} \exp(j \arg \mathbf{x}_q)$, $\arg(\mathbf{x}_q)$ 表示向量 \mathbf{x}_q 中每一元素的角度,并对于每一个向量 \mathbf{y}_q ,计算代价函数,即:

$$\beta_q = \|\mathbf{X}^* - \mathbf{y}_q \mathbf{y}_q^H\|^2, q = 1, 2, \dots, Q \quad (17)$$

则序列 $\{\beta_q\}$ 中的最小值所对应的 \mathbf{y}_q 即为满足约束条件的恒包络发射波形。

2.2 基于频域的 MIMO 雷达波形二次优化

在得到满足空域发射方向图匹配以及置零约束等条件的 MIMO 雷达恒包络发射波形后,对发射波形在频域上进行二次优化,实现在空域雷达发射方向图不变的条件下优化波形频谱,从而避免频域干扰提高雷达工作性能。

由式(6)可知, MIMO 雷达发射方向图仅与不同阵元之间发射波形的相关性有关,而与不同码元之间信号相位无关,因此改变每一码元时刻对应信号序列 $s(n)$ 的初始相位,不会对雷达发射方向图造成影响。设基于空域优化后得到的波形矩阵为 $\mathbf{S} \in \mathbb{C}^{M \times N}$,通过改变不同时刻信号序列的初始相位,得到新的波形矩阵为:

$$\tilde{\mathbf{S}} = \mathbf{S}\mathbf{A} = [e^{j\varphi_0}\mathbf{s}(0), e^{j\varphi_1}\mathbf{s}(1), \dots, e^{j\varphi_{N-1}}\mathbf{s}(N-1)] \quad (18)$$

其中, $\mathbf{A} = \text{diag}([e^{j\varphi_0}, e^{j\varphi_1}, \dots, e^{j\varphi_{N-1}}])$ 为相位变化对角矩阵,则有:

$$\tilde{\mathbf{S}}\tilde{\mathbf{S}}^H = \mathbf{S}\mathbf{A}\mathbf{A}^H\mathbf{S}^H = \mathbf{S}\mathbf{S}^H \quad (19)$$

由此可知,相位变化后的波形相关矩阵不会发生任何变化,即方向图不变。因此在不影响雷达发射方向图的基础上可以通过优化对角矩阵 \mathbf{A} ,实现波形在频域上的进一步优化。相位变化后阵元 m 发射信号序列 \tilde{s}_m 在归一化信号频带内的功率谱密度(Power Spectral Density, PSD)为^[16]:

$$\begin{aligned} \tilde{S}_m(f) &= \left| \sum_{n=0}^{N-1} \tilde{s}_m(n) e^{j\varphi_n} e^{-j2\pi fn} \right|^2 \\ &= |\mathbf{v}^H \mathbf{s}_m \odot \mathbf{F}(f)|^2 \\ &= \mathbf{v}^H (\mathbf{s}_m \odot \mathbf{F}(f)) (\mathbf{s}_m \odot \mathbf{F}(f))^H \mathbf{v} \end{aligned} \quad (20)$$

其中, $\mathbf{v} = [e^{-j\varphi_0}, e^{-j\varphi_1}, \dots, e^{-j\varphi_{N-1}}]^T$, $\mathbf{F}(f) = [1,$

$e^{-j2\pi f}, \dots, e^{-j2\pi f(N-1)}]^T$ 表示在归一化频点 f 处的傅里叶变化向量, “ \odot ” 表示 Hadamard 乘积, 信号序列 \bar{s}_m 在干扰频带 $\Omega = [f_1^i, f_2^j]$ 内的发射功率为:

$$\begin{aligned} & \int_{f_1^i}^{f_2^j} \bar{S}_m(f) df \\ &= \mathbf{v}^H \int_{f_1^i}^{f_2^j} (\mathbf{s}_m \odot \mathbf{F}(f)) (\mathbf{s}_m \odot \mathbf{F}(f))^H df \mathbf{v} = \mathbf{v}^H \mathbf{R}_J^m \mathbf{v} \end{aligned} \quad (21)$$

其中: $\mathbf{R}_J^m = \int_{f_1^i}^{f_2^j} (\mathbf{s}_m \odot \mathbf{F}(f)) (\mathbf{s}_m \odot \mathbf{F}(f))^H df$, f_1^i 表示干扰频带下边界, 即最小干扰频点; f_2^j 则表示干扰频带上边界, 即最大干扰频点。为避免频域上的干扰, MIMO 雷达发射波形在频域上的优化方程 P_4 可表示为:

$$\begin{cases} \min_{\mathbf{v}} \mathbf{v}^H \sum_{m=1}^M \mathbf{R}_J^m \mathbf{v} \\ \text{s. t.} \quad |\mathbf{V}(n)| = 1, n = 1, 2, \dots, N \end{cases} \quad (22)$$

为求解该非凸优化问题, 首先将矩阵 $\sum_{m=1}^M \mathbf{R}_J^m$ 进行特征值分解, 得到最大特征值 λ , 令 $\bar{\mathbf{R}} = \lambda \mathbf{I}_N - \sum_{m=1}^M \mathbf{R}_J^m$, 利用辅助变量 $\bar{\mathbf{R}}$ 将 P_4 等效转化为 P_5 , 即:

$$\begin{cases} \max_{\mathbf{v}} \mathbf{v}^H \bar{\mathbf{R}} \mathbf{v} \\ \text{s. t.} \quad |\mathbf{v}(n)| = 1, n = 1, 2, \dots, N \end{cases} \quad (23)$$

由于 $\bar{\mathbf{R}}$ 为 Hermitian 半正定矩阵, 且向量 \mathbf{V} 的取值范围为 $\Delta = \{\mathbf{v} \in C^N \mid |\mathbf{V}(n)| = 1, n = 0, 1, \dots, N-1\}$, 因此优化问题 P_5 为关于 \mathbf{v} 的非凸单位二次规划问题 (Unimodular Quadratic Programming, UQP), 可利用拟功率迭代算法^[17]进行有效求解。假设经过 k 次迭代后优化得到 $\mathbf{v}^{(k)}$, 则第 $k+1$ 次迭代优化方程等价于:

$$\min_{\mathbf{v}^{(k+1)} \in \Delta} \|\mathbf{v}^{(k+1)} - \bar{\mathbf{R}} \mathbf{v}^{(k)}\|_2^2 \quad (24)$$

利用拟功率算法可直接求得第 $k+1$ 次迭代的最优解, 即:

$$\mathbf{v}^{(k+1)} = e^{j \arg(\bar{\mathbf{R}} \mathbf{v}^{(k)})} \quad (25)$$

$$\begin{aligned} & \text{由于} \\ & \|\mathbf{v}^{(k+1)} - \bar{\mathbf{R}} \mathbf{v}^{(k)}\|_2^2 = \text{const} - 2\Re\{\mathbf{v}^{(k+1)H} \bar{\mathbf{R}} \mathbf{v}^{(k)}\} \end{aligned} \quad (26)$$

则 $\mathbf{v}^{(k+1)}$ 应满足 $\Re\{\mathbf{v}^{(k+1)H} \bar{\mathbf{R}} \mathbf{v}^{(k)}\}$ 最大化。若 $\mathbf{v}^{(k+1)} \neq \mathbf{v}^{(k)}$, 因为 $\bar{\mathbf{R}}$ 为 Hermitian 半正定矩阵, 则有:

$$(\mathbf{v}^{(k+1)} - \mathbf{v}^{(k)})^H \bar{\mathbf{R}} (\mathbf{v}^{(k+1)} - \mathbf{v}^{(k)}) > 0 \quad (27)$$

其中 $\Re\{\mathbf{v}^{(k+1)H} \bar{\mathbf{R}} \mathbf{v}^{(k)}\}$ 表示复数 $\mathbf{v}^{(k+1)H} \bar{\mathbf{R}} \mathbf{v}^{(k)}$ 的实部, 因此由式(27)可进一步得知:

$$\mathbf{v}^{(k+1)H} \bar{\mathbf{R}} \mathbf{v}^{(k+1)} > 2\Re\{\mathbf{v}^{(k+1)H} \bar{\mathbf{R}} \mathbf{v}^{(k)}\} - \mathbf{v}^{(k)H} \bar{\mathbf{R}} \mathbf{v}^{(k)} > \mathbf{v}^{(k)H} \bar{\mathbf{R}} \mathbf{v}^{(k)} \quad (28)$$

拟功率迭代算法的收敛性得到证明。重复上述迭代优化过程直到 $\mathbf{v}^H \bar{\mathbf{R}} \mathbf{v} \leq E_I$, E_I 为干扰频带内雷达允许最大发射功率, 停止迭代输出 \mathbf{v} 。最终经过空频域二次优化的 MIMO 雷达发射波形矩阵为:

$$\tilde{\mathbf{S}}^* = \mathbf{S} \text{Diag}(\mathbf{v}^*) \quad (29)$$

其中 $\text{Diag}(\mathbf{v}^*)$ 表示以向量 \mathbf{v}^* 构造的对角矩阵。

2.3 算法性能分析

针对干扰条件下 MIMO 雷达发射方向图优化问题, 本文提出一种基于空频域二次优化的 MIMO 雷达波形设计方法, 即首先在空域上设计与期望方向图匹配而且能够形成较宽零陷的 MIMO 雷达发射波形, 在此基础上利用阵列信号 $\mathbf{s}(n)$ 每个码元改变相同相位对应方向图不变的特性, 通过优化相位变化矩阵实现波形在频域上的优化。与文献[14]所提波形设计方法相比, 本文在利用向量方法合成具体波形时, 采用最小二乘准则使合成信号方向图逼近优化发射方向图, 从而保证了合成后的发射波形能够在空域上形成较宽的零陷; 在频域上通过优化相位变化矩阵 \mathbf{A} , 实现在不影响雷达发射方向图的条件下优化波形频谱, 从而抑制频域上的干扰。本文所提算法主要分为三部分, 基于协方差矩阵的发射方向图设计、信号合成和频谱优化, 其计算复杂度分别为 $O((M)^{3.5})$ 、 $O(Q(MN)^2)$ 、 $O(N_{\text{iter}}(N)^2)$, 其中 N_{iter} 为拟功率算法迭代次数, 相比式(13)直接对 MIMO 雷达空时序列协方差矩阵 \mathbf{X} 优化, 所提算法计算复杂度大大降低, 因此能够更好地满足雷达波形设计实时性应用的要求。

3 实验仿真

设 MIMO 雷达发射阵列为均匀线阵, 阵元间距为半波长, 阵元数目 $M = 10$, 每个阵元发射信号中心载频和信号带宽相同, 分别为 $f_0 = 10$ GHz、 $B = 10$ MHz, 雷达发射脉冲宽度 $T_p = 6.4$ μs , 发射总功率 $E = M$, 每个阵元发射基带信号码长为 $N = T_p B = 64$ 。设整个空域为 $\Theta = [-90^\circ, 90^\circ]$, 其中感兴趣的目标空域为 $\Theta_T = [-30^\circ, 30^\circ]$, 旁瓣空域为 $\Theta_S = [-90^\circ, -30^\circ] \cup [30^\circ, 90^\circ]$, 空域离散点间隔为 0.5° 。设在方向 $\theta_c = 57^\circ$ 处存在一快速移动干扰, 令导数约束 $p = 2$, 空域零陷深度 $\varepsilon = -40$ dB, 随机化实验次数 $Q = 1000$ 。

设阵元功率变化参数 $\gamma = 1$ 、 $\gamma = 1.5$ 、 $\gamma = 2$, 将本文所提通过优化协方差矩阵 $\mathbf{X} = \mathbf{I}_N \otimes \bar{\mathbf{R}}^*$ 形成带有宽零陷的 MIMO 雷达发射方向图与式(8)形

成的发射方向图进行对比,如图 1 所示。相比于式(8)形成的干扰条件下 MIMO 雷达发射方向图,本文所提方法能够通过 p 阶导数约束展宽零陷,可以较好地抵抗快速移动干扰。而且由图 1 可知,阵元功率变化参数 γ 越大,优化矩阵 \mathbf{X} 所对应的发射方向图旁瓣越低,这是因为不同阵元功率变化越大,发射波形自由度越高,因此合成方向图质量越好。此外,为进一步验证本文所提方法与直接求解式(13)所得优化方向图一致,将两者所得优化方向图进行对比,如图 2 所示。两者所得方向图完全一致,具有相同的方向图匹配误

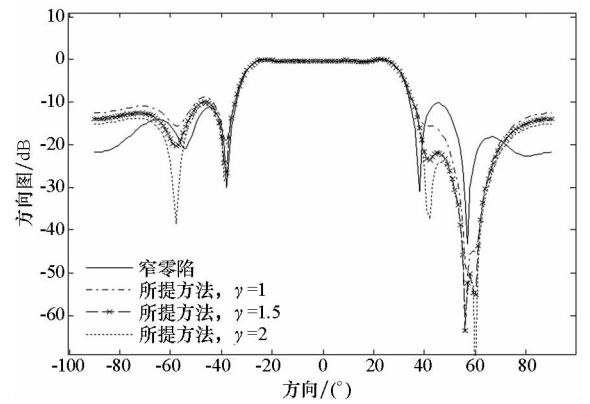


图 1 MIMO 雷达最优发射方向图

Fig. 1 Optimal transmit beampatterns of MIMO radar

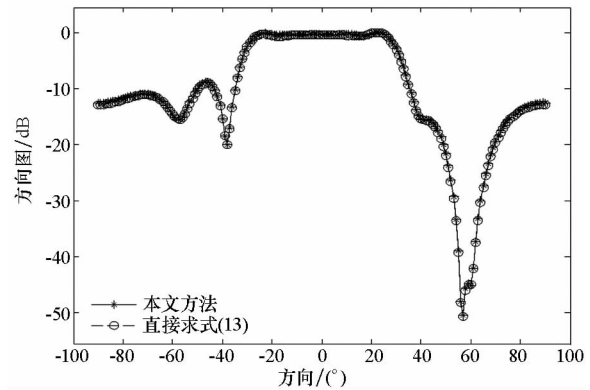


图 2 所提方法与式(13)所得优化方向图

Fig. 2 Transmit beampatterns by the proposed method and equation (13)

差,但是直接求解式(13)计算复杂度为 $O((MN)^{3.5})$,而本文所提方法求解 P_3 计算复杂度仅为 $O(M^{3.5})$,在发射信号码长较大时,本文所提在空域上的波形优化方法比直接求解式(13)计算效率大幅提升。

为更好地分析基于随机向量合成方法和文献[15]所提循环算法(Cyclic Algorithm, CA)在干扰条件下合成波形的质量,令 $\gamma = 1$,基于两种

恒包络波形设计方法所得 MIMO 雷达发射方向图如图 3 所示。由图 3 可以直观看出,CA 方法所得波形不能保证雷达发射方向图在干扰方向上形成满足条件的零陷,这是因为该循环算法在合成信号矩阵 \mathbf{S} 时,仅以最小二乘准则逼近矩阵 $\mathbf{R}^{1/2}\mathbf{U}$,而没有考虑零陷约束。相比于 CA 算法,基于随机向量合成的波形设计方法以最小二乘准则逼近最优协方差矩阵,因此合成的波形不仅能够较好地匹配最优协方差矩阵 \mathbf{X} 所对应的发射方向图,而且能够保证在干扰方向形成满足一定宽度和深度的零陷。图 4 则表示了在不同功率变化参数 γ 优化情况下,本文所提方法合成的信号矩阵 \mathbf{S} 所对应的每个阵元的发射功率分配情况。虽然优化后的发射波形每个阵元发射功率不同,但是由于优化过程中定义发射空时序列协方差矩阵 $\mathbf{X} = \mathbf{I}_N \otimes \mathbf{R}$,因此每个阵元发射波形仍然保持恒包络特性。

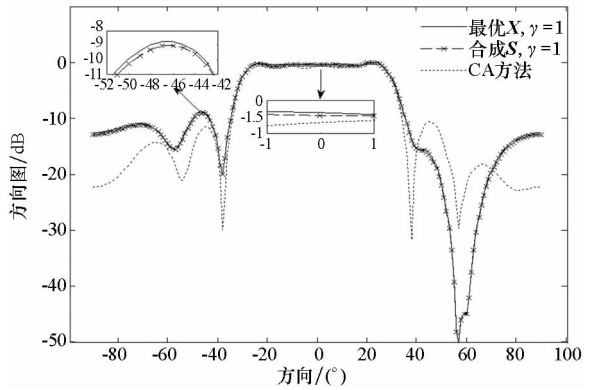


图 3 不同波形设计方法形成的 MIMO 雷达发射方向图
Fig. 3 MIMO radar transmit beampatterns synthesized by different waveform design methods

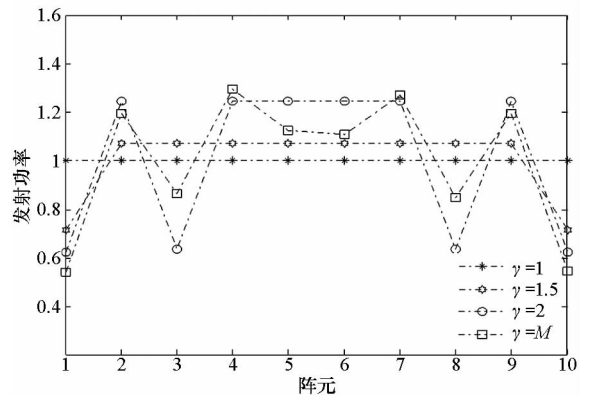


图 4 不同 γ 条件下阵元发射功率

Fig. 4 Element transmit power with different γ

本文所提基于空频域联合优化 MIMO 雷达波形设计方法中,在利用随机向量合成方法对优化协方差矩阵分解得到具体发射波形时,随机化向

量实验次数直接决定所得波形质量。定义所得波形协方差矩阵 $\hat{\mathbf{X}}$ 与优化矩阵 \mathbf{X}^* 之间均方误差 (Mean-Squared Error, MSE) 为:

$$MSE = \|\hat{\mathbf{X}} - \mathbf{X}^*\|_2^2 \quad (30)$$

在 $\gamma=1$ 、蒙特卡洛次数为 100 的条件下, 波形合成均方误差随随机化实验次数的变化情况如图 5 所示。由图 5 可知, 随着实验次数的增加波形合成误差变小, 当 $Q \geq 800$ 时, 合成波形误差几乎不变, 因此当实验次数足够大时, 合成波形能够较好地匹配优化协方差矩阵, 从而保证发射方向图的质量。

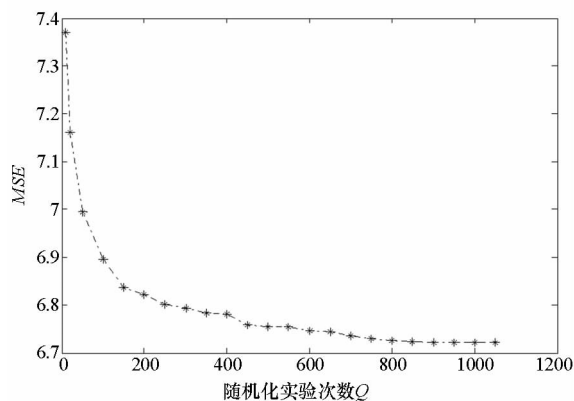


图 5 均方误差随实验次数变化情况

Fig. 5 MSE versus the number of experiments

设在频域上干扰信号归一化带宽为 $\Omega = [0.4, 0.5]$, 雷达发射信号在干扰频带内允许的最大发射功率为 $E_t = -40$ dB。在 $\gamma=1$ 条件下, 将本文所提空频域二次优化方法与式 (13) 仅在空域进行优化所得波形的功率谱密度进行对比, 后者得到的优化波形功率谱在频域上任意分布, 无法有效抵抗频域上的干扰, 如图 6 所示。而本文所提基于频域二次优化后的 MIMO 雷达发射波形功率谱如图 7 所示。在保证 MIMO 雷达空域发射方向图不变的情况下, 通过优化发射波形初始

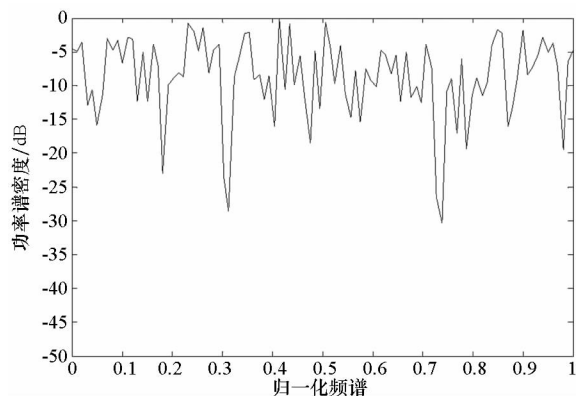


图 6 基于空域优化后的波形功率谱

Fig. 6 PSD of optimized waveforms via spatial optimization

相位矩阵 \mathbf{A} , 可以较好地控制波形频谱在干扰频带内总的发射功率, 从而将雷达发射波形规避干扰带宽, 实现频域上的干扰抑制。

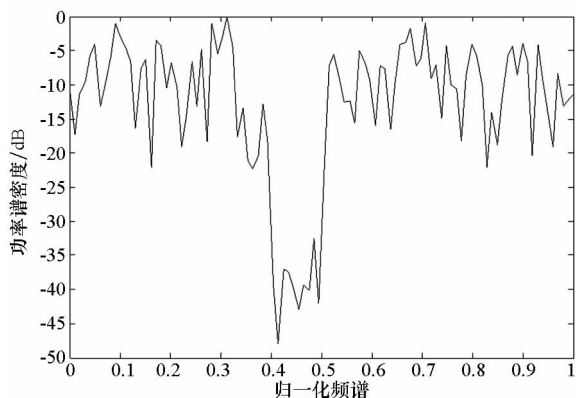


图 7 基于频域二次优化后的波形功率谱

Fig. 7 PSD of optimized waveforms via the second spectral optimization

4 结论

本文针对一般 MIMO 雷达波形设计方法不能在匹配期望发射方向图的同时抑制来自空域和频域的干扰问题, 提出一种干扰条件下基于空频域二次优化的 MIMO 雷达波形设计方法。该方法首先利用 MIMO 雷达发射方向图表达式, 将空域上方向图优化问题转化为关于雷达空时序列协方差矩阵 \mathbf{X} 的优化问题, 并利用 MIMO 雷达发射方向图仅与不同阵元之间发射波形相关性有关这一特性, 进一步降低空域上波形优化计算复杂度, 针对空域上快速移动的干扰, 通过 p 阶导数约束实现零陷展宽; 通过优化得到最优协方差矩阵 \mathbf{X}^* 后, 利用随机向量方法通过最小二乘准则逼近最优发射方向图来合成恒包络发射波形, 不仅能够较好地匹配最优协方差矩阵 \mathbf{X}^* 所对应的发射方向图, 而且能够保证在干扰方向形成满足一定宽度和深度的零陷; 最后在空域优化得到的发射波形基础上, 利用改变不同时刻信号序列的初始相位雷达发射方向图不变的特性, 通过拟功率算法对相位变化矩阵 \mathbf{A} 进行优化, 从而实现 MIMO 雷达发射波形在频域上的二次优化。实验仿真证明了所提方法在方向图设计和空频域干扰抑制方面的有效性。

参考文献 (References)

- [1] Haimovich A M, Blum R S, Cimini L J. MIMO radar with widely separated antennas [J]. IEEE Signal Processing Magazine, 2008, 25(1): 116-129.
- [2] 许红波, 王怀军, 陆珉, 等. 一种新的 MIMO 雷达 DOA 估计方法[J]. 国防科技大学学报, 2009, 31(3): 92-96.

- XU Hongbo, WANG Huaijun, LU Min, et al. A new algorithm on estimation of DOA using MIMO radar [J]. Journal of National University of Defense Technology, 2009, 31(3): 92–96. (in Chinese)
- [3] Li J, Stocia P. MIMO radar with colocated antennas [J]. IEEE Signal Processing Magazine, 2007, 24(5): 106–114.
- [4] Fuhrmann D R, Antonio G S. Transmit beamforming for MIMO radar systems using signal cross-correlation [J]. IEEE Transactions on Aerospace and Electronic Systems, 2008, 44(1): 171–186.
- [5] Stocia P, Li J, Yao X. On probing signal design for MIMO radar[J]. IEEE Transactions on Signal Processing, 2007, 55(8): 4151–4161.
- [6] Pandey N, Roy L P. Convex optimization based transmit beampattern synthesis for MIMO radar [J]. Electronic Letters, 2016, 52(9): 761–763.
- [7] Ahmed S, Alouini M S. MIMO radar transmit beampattern design without synthesising the covariance matrix[J]. IEEE Transactions on Signal Processing, 2014, 62(9): 2278–2289.
- [8] Imani S, Ghorashi S A, Bolhasani M. SINR maximization on colocated MIMO radars using transmit covariance matrix[J]. Signal Processing, 2016, 119: 128–135.
- [9] Khabbazihasmenj A, Hassanien A, Vorobyov S, et al. Efficient transmit beamspace design for search-free based DOA estimation in MIMO radar[J]. IEEE Transactions on Signal Processing, 2014, 62(6): 1490–1500.
- [10] Friedlander B. On transmit beamforming for MIMO radar[J]. IEEE Transactions on Aerospace and Electronic Systems, 2012, 48(4): 3376–3388.
- [11] Li Y Z, Vorobyov S A, Koivunen V. Ambiguity function of the transmit beamspace-based MIMO radar [J]. IEEE Transactions on Signal Processing, 2015, 63(17): 4445–4457.
- [12] Hua G, Abeysekera S S. MIMO radar transmit beampattern design with ripple and transition band control [J]. IEEE Transactions on Signal Processing, 2013, 61(11): 2963–2974.
- [13] Gong P C, Shao Z H, Tu G P, et al. Transmit beampattern design based on convex optimization for MIMO radar systems[J]. Signal Processing, 2014, 94: 195–201.
- [14] Stocia P, Li J, Zhu X M. Waveform synthesis for diversity-based transmit beampattern design [J]. IEEE Transactions on Signal Processing, 2008, 56(6): 2593–2598.
- [15] Tang B, Tang J. Joint design of transmit waveforms and receive filters for MIMO radar space time adaptive processing [J]. IEEE Transactions on Signal Processing, 2016, 64(18): 4707–4722.
- [16] Aubry A, De Maio A, Huang Y, et al. A new radar waveform design algorithm with improved feasibility for spectral coexistence [J]. IEEE Transactions on Aerospace and Electronic Systems, 2015, 51(2): 1029–1038.
- [17] Soltanalian M, Stoica P. Designing unimodular codes via quadratic optimization [J]. IEEE Transactions on Signal Processing, 2014, 62(5): 1221–1234.

全极化雷达的多任务压缩感知目标识别方法*

翟庆林¹, 刘盛启², 胡杰民¹, 占荣辉¹

(1. 国防科技大学 电子科学与工程学院, 湖南 长沙 410073; 2. 中国人民解放军 31011 部队, 北京 100091)

摘要:为有效利用全极化雷达高分辨距离像(High Resolution Range Profile, HRRP)的丰富特征信息和全极化样本中各单极化 HRRP 均对应于相同目标姿态的特性,提出一种基于多任务压缩感知的全极化雷达目标识别方法。该方法约束在不同极化字典中选择来自相同角域的字典原子对相应极化方式下的 HRRP 进行表示,可以有效利用不同极化 HRRP 之间的相关信息提高目标识别性能。基于电磁散射数据对所提出的方法进行了测试,实验结果证明了方法的有效性。

关键词:雷达目标识别;全极化高分辨距离像;多任务压缩感知

中图分类号:TN919 **文献标志码:**A **文章编号:**1001-2486(2017)03-144-07

Full-polarization radar target recognition of multitask compressive sensing

ZHAI Qinglin¹, LIU Shengqi², HU Jiemin¹, ZHAN Ronghui¹

(1. College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China;
2. The PLA Unit 31011, Beijing 100091, China)

Abstract: To efficiently utilize the information which can be extracted for target recognition and the character that different polarization channels characterize the same structure signature of a target using different polarization modes to boost recognition performance, a method for full-polarization HRRP recognition based on multitask compressive sensing was proposed. Each single-polarization HRRP was represented by the atoms adaptively selected from its associated dictionary, and the atoms derived from different dictionaries corresponded to the same index set. Compared with the conventional methods, the proposed method has the significant advantage of exploiting the correlation among single-polarization HRRPs to enhance recognition performance. Experiments were carried out on simulated data, and the results demonstrate the efficiency of the proposed method.

Key words: radar target recognition; full-polarization HRRP; multitask compressive sensing

宽带全极化雷达综合了高分辨技术和全极化测量的优点,为目标识别提供了更为丰富的特征信息。文献[1]指出,极化与高分辨技术的结合大概是最有希望解决目标识别问题的研究方向;如何有效地利用宽带多极化信息也成了雷达目标识别领域的研究热点^[2-9]。高分辨信息与极化信息相结合在雷达目标识别领域的应用主要有两条途径:一是基于宽带雷达目标回波极化特征的识别,例如基于极化散射矩阵(Polarization Scattering Matrix, PSM)的识别方法^[2-3]和基于目标分解(Target Decomposition, TD)的识别方法^[4];二是将极化测量与高分辨成像相结合的识别方法,例如基于极化高分辨距离像(High Resolution Range Profile, HRRP)的识别^[5-6]以及基于极化合成孔

径雷达/逆合成孔径雷达(Synthetic Aperture Radar/Inverse SAR, SAR/ISAR)图像的识别^[2,7-8]等。

本文主要关注基于全极化 HRRP 的目标识别问题。多极化 HRRP 目标识别主要基于数据融合的方式进行,包括特征层以及决策层的融合方法。特征层融合方法将 HRRP 与极化信息按照某种规则结合起来,再构造相应的分类器进行目标类型判决。决策层融合方法首先对各极化 HRRP 独立进行识别,再选择适当的算法融合多分类器输出获得最终的目标类型判定^[7,9-10]。决策层融合方法在识别过程中没有考虑不同极化 HRRP 之间的相互关系,显然无法获得最优的目标识别性能。

* 收稿日期:2016-05-09
基金项目:国家自然科学基金资助项目(61471370,61401479)
作者简介:翟庆林(1980—),男,山东烟台人,副教授,博士,E-mail:qinglinzhai@139.com

宽带全极化雷达可以同时获得目标的全极化 HRRP,在成像瞬间目标姿态近似不变,目标散射结构也保持不变。故不同极化 HRRP 描述的是,相同的目标散射结构、各极化分量间应具有一定的相关性。由同一姿态角下目标归一化的多极化 HRRP 可以看出(如图 1 所示),不同极化(HH/HV/VH/VV,其中 H 表示水平极化,V 表示垂直极化)HRRP 同一距离单元的回波幅度存在明显的差异,但其散射中心位置分布基本一致,说明不同极化 HRRP 间确实存在一定的相关性。因此,本文提出了一种基于多任务压缩感知的全极化 HRRP 目标识别方法。

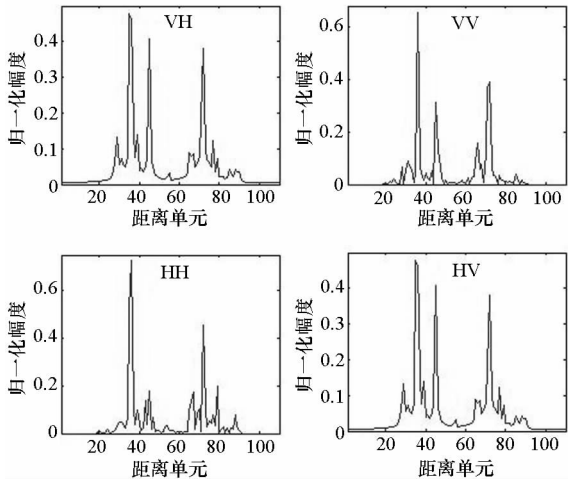


图 1 全极化 HRRP 示意

Fig. 1 Illustration of full-polarization HRRP

1 多任务压缩感知

多任务压缩感知 (MultiTask Compressive Sensing, MTCS)^[11] 针对具有相关性的多个压缩观测向量的重构问题,从贝叶斯分析的角度出发求解稀疏系数的最大后验概率估计。MTCS 将机器学习领域的多任务学习方法引入稀疏系数重构过程,通过在多任务间共享信息建立多任务之间的联系,可以有效利用多观测间的统计相关性进行信号重构。

假设 L 个统计相关的观测向量 $\{y_i | i = 1, \dots, L\}$ 可表示为:

$$y_i = A_i x_i + n_i \quad (1)$$

式中, $A_i \in \mathbf{R}^{N_i \times N}$ ($N_i < N$) 为第 i 个传感矩阵, $x_i \in \mathbf{R}^N$ 为 K_i 阶稀疏信号(即 x_i 中仅有 K_i 个元素非零), n_i 为零均值、精度 ξ_0 (方差的倒数)未知的独立同分布高斯变量。

y_i 的似然函数可以表示为:

$$p(y_i | x_i, \xi_0) = (2\pi/\xi_0)^{-N_i/2} \exp\left(-\frac{\xi_0}{2} \|y_i - A_i x_i\|_2^2\right) \quad (2)$$

MTCS 假设参数 $\{x_i | i = 1, \dots, L\}$ 均服从相同的先验分布,从而建立起多任务间的统计相关性联系。MTCS 假定 x_i 的先验分布为:

$$p(x_i | \xi) = \prod_{j=1}^N N(x_{i,j} | 0, \xi_j^{-1}) \quad (3)$$

式中, $x_{i,j}$ 表示 x_i 的第 j 个元素, ξ_j 为高斯分布的精度。参数 $\xi = [\xi_1 \dots \xi_N]^T$ 在多任务间是固定的,由 L 个观测共同求解获得。

ξ 及 ξ_0 均采用 Gamma 先验:

$$p(\xi_0 | a, b) = Ga(\xi_0 | a, b) \quad (4)$$

$$p(\xi | c, d) = \prod_{j=1}^N Ga(\xi_j | c, d) \quad (5)$$

当 ξ 已知时, x_i 的后验概率密度函数为:

$$p(x_i | y_i, \xi) = \int p(x_i | y_i, \xi, \xi_0) p(\xi_0 | a, b) d\xi_0 \\ = [1 + (x_i - \mu_i)^T \Sigma_i^{-1} (x_i - \mu_i) / 2b]^{-(a+N/2)} \cdot \frac{\Gamma(a + N/2)}{\Gamma(a) (2\pi b)^{N/2} |\Sigma_i^{-1}|^{1/2}} \quad (6)$$

均值和协方差矩阵分别为

$$\mu_i = \Sigma_i A_i^T y_i \quad (7)$$

$$\Sigma_i = (A_i^T A_i + \Lambda)^{-1} \quad (8)$$

其中 $\Lambda = \text{diag}(\xi_1, \xi_2, \dots, \xi_N)$ 。

ξ 可通过最大化式(9)进行估计:

$$L(\xi) = \sum_{i=1}^L \lg p(y_i | \xi) \\ = -\frac{1}{2} \sum_{i=1}^L (N_i + 2a) \lg(y_i^T B_i^{-1} y_i + b) - \frac{1}{4} \sum_{i=1}^L \lg |B_i| + \text{const} \quad (9)$$

式中, $B_i = I + A_i \Lambda^{-1} A_i^T$ 。

考虑到 ξ_j 与 $L(\xi)$ 的依赖关系,可将 B_i 分解为 $B_i = B_{i,-j} + \xi_j^{-1} A_{i,j} A_{i,j}^T$, 其中 $B_{i,-j} = I + \sum_{k \neq j} \xi_k^{-1} A_{i,k} A_{i,k}^T$ 。则 $L(\xi)$ 可表示为:

$$L(\xi) = L(\xi_{-j}) - \frac{1}{2} \lg(1 + s_{i,j}/\xi_j) - \frac{1}{2} \sum_{i=1}^L (N_i + 2a) \lg\left(1 - \frac{q_{i,j}^2/g_{i,j}}{\xi_j + s_{i,j}}\right) \quad (10)$$

式中, ξ_{-j} 表示去除 ξ 中第 j 个元素后余下的向量, 并且:

$$\begin{cases} s_{i,j} = A_{i,j}^T B_{i,-j}^{-1} A_{i,j} \\ q_{i,j} = A_{i,j}^T B_{i,-j}^{-1} y_i \\ g_{i,j} = y_i^T B_{i,-j}^{-1} y_i + 2b \end{cases} \quad (11)$$

对 $L(\xi)$ 求 ξ_j 的偏导数并令其值为零。一般

情况下有 $\xi_j \ll s_{i,j}$, 则 ξ_j 可近似表示为:

$$\xi_j \approx \begin{cases} \frac{L}{E_j}, & E_j > 0 \\ \infty, & \text{otherwise} \end{cases} \quad (12)$$

式中, $E_j = \sum_{i=1}^L \frac{(N_i + 2a) q_{i,j}^2 / g_{i,j} - s_{i,j}}{s_{i,j} (s_{i,j} - q_{i,j}^2 / g_{i,j})}$.

通过式(12)可以控制 $A_{i,j}$ 是否用于测试向量的稀疏表示。为避免矩阵求逆运算, $s_{i,j}, q_{i,j}, g_{i,j}$ 可通过式(13)计算:

$$\begin{cases} s_{i,j} = \frac{\xi_j S_{i,j}}{\xi_j - S_{i,j}} \\ q_{i,j} = \frac{\xi_j Q_{i,j}}{\xi_j - Q_{i,j}} \\ g_{i,j} = G_i + \frac{Q_{i,j}^2}{\xi_j - S_{i,j}} \end{cases} \quad (13)$$

其中 $S_{i,j}, Q_{i,j}$ 及 G_i 定义为:

$$\begin{cases} S_{i,j} = A_{i,j}^T A_{i,j} - A_{i,j}^T A_i \Sigma_i A_i^T A_{i,j} \\ Q_{i,j} = A_{i,j}^T y_i - A_{i,j}^T A_i \Sigma_i A_i^T y_i \\ G_i = y_i^T y_i - y_i^T A_i \Sigma_i A_i^T y_i + 2b \end{cases} \quad (14)$$

这里 A_i 和 Σ_i 仅包括当前活动的基向量。

注意到 ξ 是 μ_i 和 Σ_i 的函数, 而 μ_i 和 Σ_i 又是 ξ 的函数, 因此可以利用迭代算法在式(7)、式(8)、式(12)之间进行迭代求解。满足收敛条件后, 即可利用式(7)估计稀疏系数。

2 MTCS 的全极化 HRRP 目标识别

利用宽带全极化雷达进行一维成像, 可以获得目标 4 种极化组合方式下的 HRRP。设 P 表示发送电磁波极化方式, Q 表示接收电磁波极化方式, 则全极化 HRRP 可表示为:

$$\{y^{PQ} | P, Q = H, V\} \quad (15)$$

假设训练阶段模板库中存在 C 类目标, 第 c ($c = 1, \dots, C$) 类目标在 PQ 极化方式下的训练样本集为 A_c^{PQ} 。文献[12-13]指出, 当训练样本充足时, 测试样本可以用来自同一目标的训练样本进行线性表示:

$$y^{PQ} = A_c^{PQ} x_c^{PQ} \quad (16)$$

式中, y^{PQ} 假设为来自第 c 类目标 PQ 极化方式下的 HRRP 测试样本, x_c^{PQ} 为 y^{PQ} 在 A_c^{PQ} 上对应的线性表示系数。由于实际应用中测试样本的类别是未知的, y^{PQ} 应该用训练集中所有目标类别的训练样本进行表示, 即:

$$y^{PQ} = A^{PQ} x^{PQ}, \quad P, Q = \{H, V\} \quad (17)$$

式中, $A^{PQ} = [A_1^{PQ} \dots A_C^{PQ}]$ 表示所有目标的训练样本集, x^{PQ} 为 y^{PQ} 在字典 A^{PQ} 上对应的线性表示系数。理想情况下, x^{PQ} 的非零值应全部对应于第 c

类目标的训练样本, 即 $x^{PQ} = [\mathbf{0}^T \dots (x_c^{PQ})^T \dots \mathbf{0}^T]^T$ 。

由全极化 HRRP 特性可知, 全极化样本中 4 个单极化分量均对应于相同的目标姿态, 并且不同极化分量间具有一定的相关性。为有效利用这两个方面的信息提高雷达目标识别性能, 引入 MTCS 求解式(17)的稀疏系数重构问题。为此, 首先需要构造满足要求的全极化字典。本文中, 过完备字典的构建基于训练数据进行, 字典构造方法如图 2 所示。为克服 HRRP 的姿态敏感性, 需要对全方位的训练数据进行分帧处理, 再对帧内训练样本进行特征提取构造相应极化方式下的过完备字典。

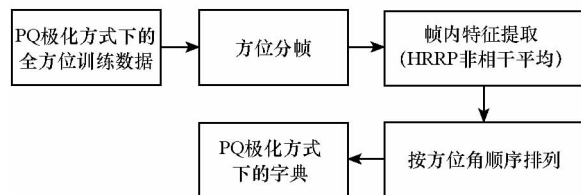


图 2 全极化字典构造方法

Fig. 2. Flowchart of dictionary construction

满足字典原子构造要求后, 利用 MTCS 方法可以在噪声环境下精确估计稀疏系数。获得稀疏系数的估计值后, 还需要设计相应的分类器将测试数据划分为特定的目标类别。本文利用总重构误差最小准则判决目标类别, 分类准则为:

$$\hat{c} = \underset{c \in \{1, \dots, C\}}{\operatorname{argmin}} \sum_P \sum_Q \|y^{PQ} - A^{PQ} \sigma_c(x^{PQ})\|_2^2 \quad (18)$$

式中, $\sigma_c(x^{PQ})$ 表示保留 x^{PQ} 中对应第 c 类训练样本的系数元素, 并将其他元素置零。

综上所述, 基于 MTCS 的全极化 HRRP 目标识别算法 (MTCS-FP) 如算法 1 所示。

算法 1 MTCS-FP 算法

Alg. 1 MTCS-FP algorithm

输入: 全极化 HRRP 测试样本 $\{y^{PQ}\}$, 全极化字典 $\{A^{PQ}\}$, $P, Q \in \{H, V\}$;

输出: 测试样本所属的目标类别估计 \hat{c} ;

1) 利用 MTCS 估计 $\{y^{PQ}\}$ 在对应极化字典上的稀疏表示系数 $\{x^{PQ}\}$;

2) 重构全极化测试样本: $\hat{y}_c^{PQ} = A^{PQ} \sigma_c(x^{PQ})$;

3) 计算重构误差: $e_c = \sum_P \sum_Q \|y^{PQ} - \hat{y}_c^{PQ}\|_2$;

4) 估计测试样本所属类别: $\hat{c} = \underset{c \in \{1, \dots, C\}}{\operatorname{argmin}} e_c$ 。

3 实验结果及分析

实验中将 MTCS-FP 算法与其他 4 种现有算

法进行了对比,包括相关匹配多数投票融合 (Matching Score Majority of Voting, MSMV) 方法^[9]、稀疏表示分类 (Sparse Representation Classifier, SRC) 方法^[12]、联合稀疏表示分类 (Joint Sparse Representation Classifier, JSRC) 方法^[14]以及联合动态稀疏表示分类 (Joint Dynamic SRC, JDSRC) 方法^[15]。

目标识别实验基于电磁仿真数据进行。目标特征库中含有 4 类地面目标,目标 CAD 模型如图 3 所示^[16]。基于 CAD 模型,将表 1 所示的电磁参数输入电磁计算软件获得目标全方位角度下的电磁散射数据。电磁计算在不同的俯仰角下进行,27°俯仰角观测数据用于训练,30°用于测试。对目标同一观测角度下的频率采样数据通过 IFFT 合成距离像。为消除 HRRP 的姿态、平移以及强度敏感性的影响,需要对距离像进行相应的

预处理^[17]。由于电磁仿真数据为类转台数据,不需要进行平移对准,实际应用中可以利用包络对齐技术消除 HRRP 的平移敏感性。为了松弛 HRRP 的姿态敏感性,按照散射中心不发生越距离单元走动 (Migration Through Resolution Cell, MTRC)^[18]的约束条件,训练阶段按 3°方位间隔对全方位回波数据进行角域划分并取各角域内 HRRP 的相干平均作为训练模板,测试阶段取 1°方位角范围内 HRRP 的非相干平均作为测试样本对各算法进行性能测试。为提供一个公平的比较,SRC,JSRC,JDSRC 以及 MTCS-FP 方法均采用相同的特征字典。图 3 中 CAD 模型的右边为该模型对应目标在同一观测角度、HH 极化方式下的归一化 HRRP 示意图。由图可以看出,各目标 HRRP 在形状、分布上存在差异,这些差异信息就构成了利用 HRRP 进行目标识别的物理基础。

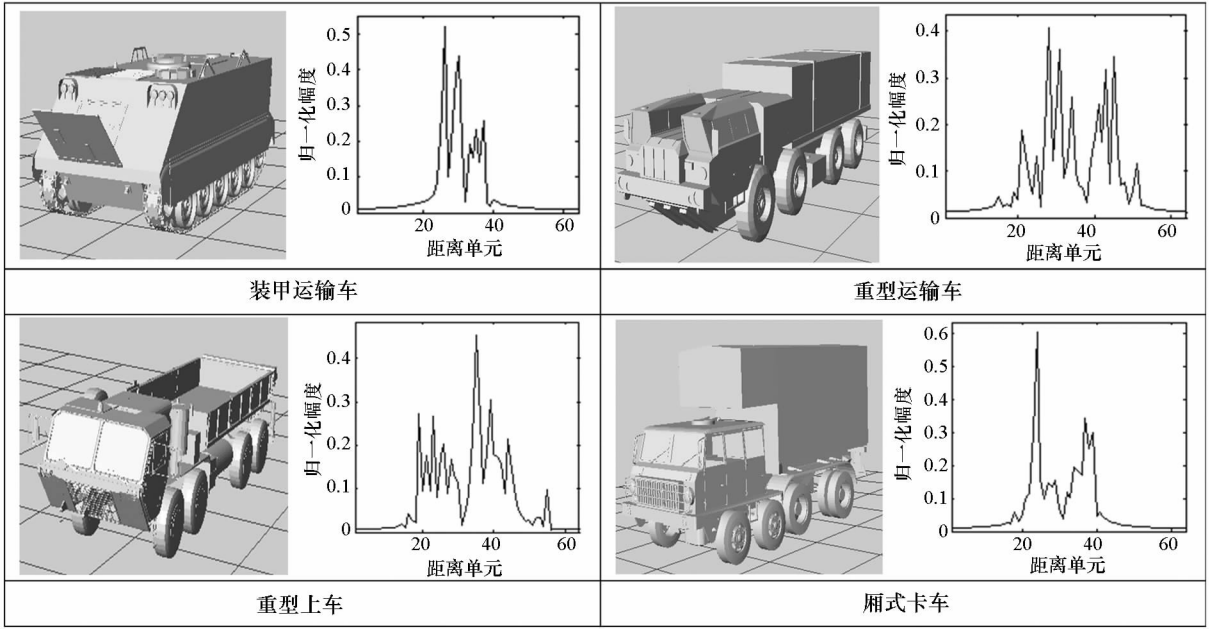


图3 地面目标模型及 HRRP 特征

Fig. 3 Simulated vehicle targets and HRRP characteristics

表 1 电磁仿真参数

Tab. 1 Electromagnetic simulation parameters

中心频率	带宽	频率采样点数	方位角	俯仰角	极化方式
10 GHz	500 MHz	128	0° ~ 360°, 0.1°间隔	27°, 30°	HH/HV/VH/VV

为测试算法在噪声条件下的识别性能,实验中在测试 HRRP 时加入零均值高斯白噪声对算法进行测试。信噪比 (Signal-to-Noise Ratio, SNR) 定义为:

$$SNR = 10 \lg \left(\frac{\sum_{l=1}^L p_l}{L \sigma^2} \right) \quad (19)$$

式中, σ^2 为高斯白噪声方差, p_l 为 HRRP 第 l 个距离单元功率, L 为距离像长度。本文中噪声环境下的目标识别结果均由 100 次蒙特卡洛实验得出,后面不再一一说明。

实验中首先对比单极化与全极化 HRRP 的目标识别性能,验证利用多极化数据对提高目标识别性能的作用,实验结果如图 4 所示。由图 4

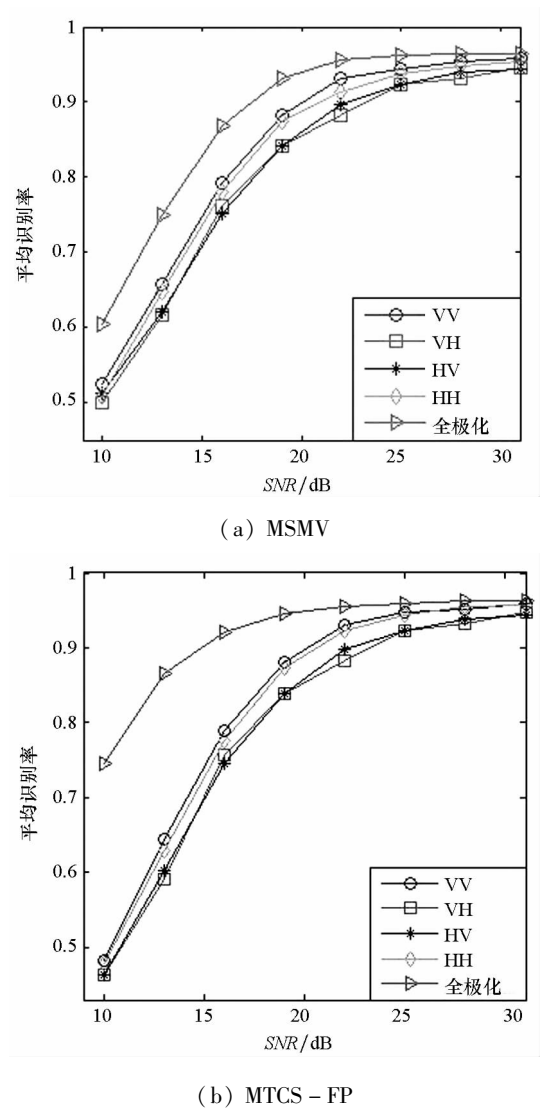


图 4 单极化 HRRP 识别结果与全极化 HRRP 识别结果对比

Fig. 4 Recognition results of single-polarization HRRPs and full-polarization HRRPs

可看出,利用全极化 HRRP 进行目标识别确实可以获得比单极化更好的目标识别性能。

接下来对不同算法的全极化 HRRP 目标识别性能进行比较。图 5 给出了各算法在不同噪声观测条件下的识别结果。从图中可以看出,MTCS - FP 在不同 SNR 条件下均具有最佳的目标识别性能。

下面对 MTCS 及 JSRC 的算法复杂度进行分析。MTCS 与 JSRC 的计算复杂度分别为 $\mathcal{O}(LNm^2)$ 、 $\mathcal{O}(NLN_i + 2NTLN_i)$, 其中 m 为稀疏度, T 为平均迭代次数。一般情况下有 $m \ll N_i$, 故 MTCS 算法的计算效率通常要高于 JSRC 方法。为获得直观的比较,对两种方法完成 4 类目标分类(共 1440 个全极化测试样本)所耗费的总时间进行了对比(见表 2)。分类实验基于相同的计算

平台进行。由表 2 的分类时间对比可以看出,MTCS - FP 的计算效率优于 JSRC 方法。

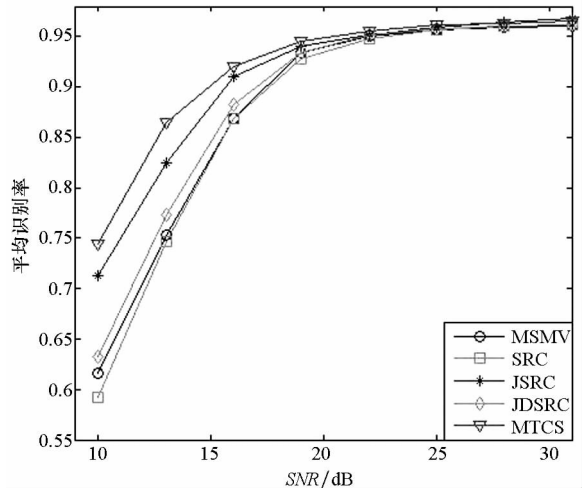


图 5 不同方法的全极化 HRRP 目标识别结果

Fig. 5 Recognition performance of HRRP using different algorithms

表 2 MTCS - FP 与 JSRC 耗时对比
Tab. 2 Runtimes of MTCS - FP and JSRC

	MTCS - FP	JSRC
耗时/s	24. 27	115. 59

为了对测试样本进行完全表示,不同极化方式下的特征字典均包含了该极化方式下目标全方位的特征信息,字典原子则对应了目标在不同姿态下的散射特性。JSRC 和 MTCS - FP 约束不同极化分量对应的稀疏系数在原子级具有相同的稀疏模式,也就是选取了目标相同姿态的字典原子对不同极化 HRRP 分量进行表示,从而可以有效利用全极化 HRRP 包含的先验信息用于识别。

最后,为直观说明 MTCS - FP 具有最佳识别性能的原因,比较了两组典型测试样本利用不同方法求解得到的稀疏系数及重构误差,如图 6 所示。稀疏系数子图的横坐标为字典原子序号,1 ~ 120,121 ~ 240,241 ~ 360,361 ~ 480 分别对应装甲运输车、重型运输车、重型卡车以及厢式卡车。为了便于说明,分别记为目标 1 ~ 4。重构误差子图的横坐标则分别对应这 4 类目标。图中的两组测试数据(记为 G1,G2)均来自第 1 类目标,每 1 行的子图为利用同一识别方法从两组测试数据求解得到的实验结果,1 ~ 4 行分别对应 SRC,JDSRC,JSRC 以及 MTCS - FP 方法。每 1 列的实验结果对应 1 组测试数据。图 6 中第 1 列的子图为 G1 的实验结果,由重构误差子图可以看出,4 种方法均在目标 1 上具有最小的重构误差,

MTCS-FP在正确的目标类别上重构误差最小。图 6 中第 2 列的子图对应从 G2 样本获得的实验结果。从图中可见, SRC, JDSRC 与 JSRC 方法均在目标 4 上具有最小的重构误差, 即这 3 种方法均产生了误判, 而 MTCS-FP 在目标 1 上重构误差最小, 说明 MTCS-FP 仍获得了正确的目标判决。通过对比这 4 种方法得到的稀疏系数可以找到产生这些差异的原因。SRC 得到的稀疏系数分布无明显规律, 非零元素杂乱分布; JDSRC 从

不同极化分量得到的稀疏系数在相同目标类别的训练集上具有相同的非零元素个数, 但类别内稀疏系数为非零值的位置并不相同; JSRC 与 MTCS-FP 方法得到的稀疏系数均在相同的位置取得非零值。MTCS-FP 可以同时利用全极化 HRRP 包含的 3 个层次的先验信息, 并且 MTCS-FP 得到稀疏系数的最大后验概率估计过程具有潜在的降噪效果, 因此 MTCS-FP 具有最优的目标识别性能。

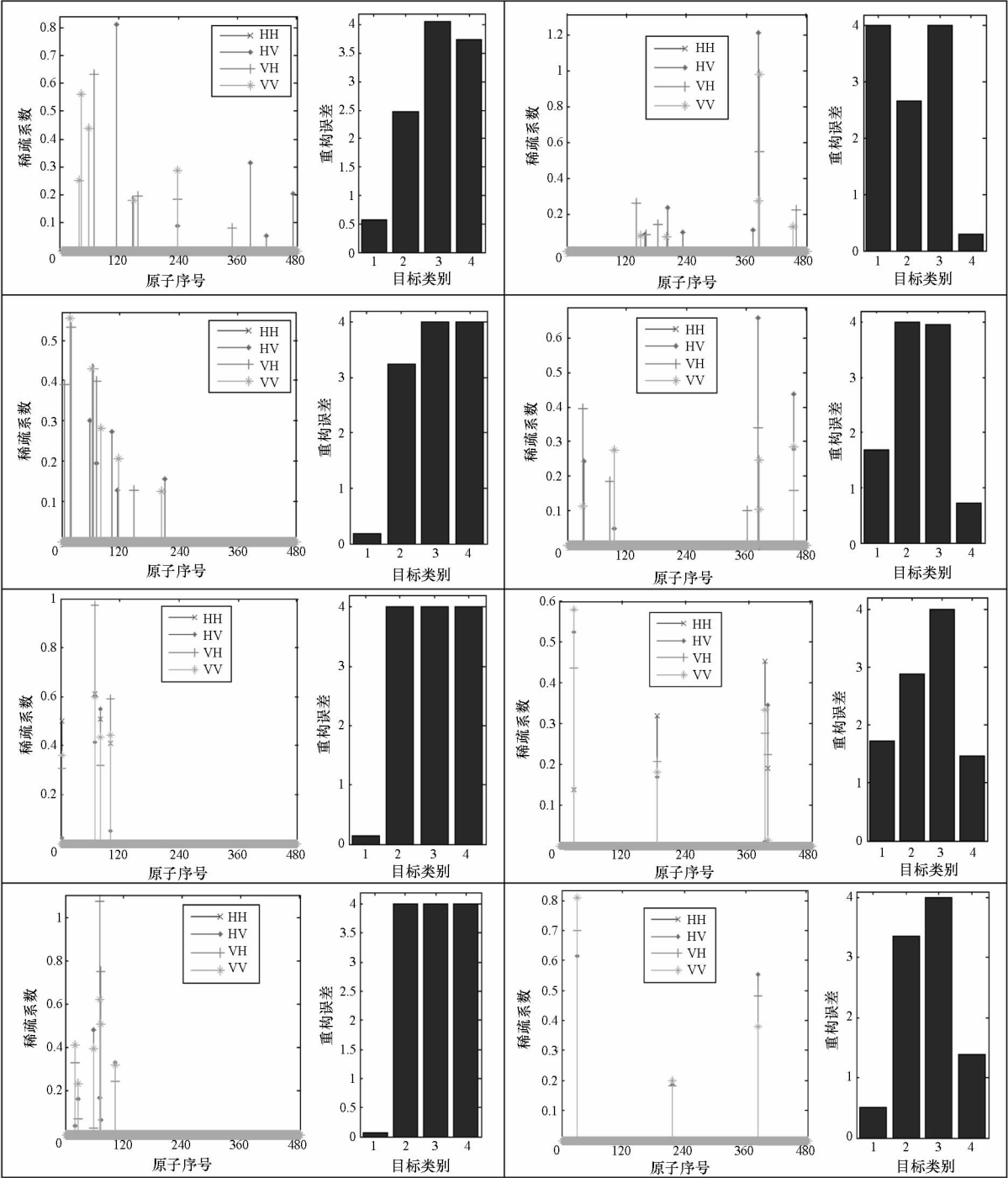


图 6 不同方法求解得到的稀疏系数及重构误差

Fig. 6 Sparse representation vectors and class-wise reconstruction errors under different algorithm

4 结论

本文针对全极化 HRRP 目标识别问题,提出了一种基于 MTCS 的全极化 HRRP 目标识别方法。该方法约束不同极化 HRRP 对应的稀疏系数在相同的稀疏模式下为非零值,可以充分利用各单极化 HRRP 对应目标相同姿态及 HRRP 间具有相关性的先验信息提高目标识别性能。实验结果表明,该方法可以提高雷达目标识别性能,并且在噪声环境下具有较好的鲁棒性。

参考文献 (References)

[1] Eaves J L, Reedy E K. Principles of modern radar [M]. New York, US:Van Nostrand Reinhold, 1987.

[2] Chen C T, Chen K S, Lee J S. The use of fully polarimetric information for the fuzzy neural classification of SAR images [J]. IEEE Transactions on Geoscience and Remote Sensing, 2003, 41(9): 2089 – 2100.

[3] Titin-Schnaider C. Characterization, recognition of bistatic polarimetric mechanisms [J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 51(3): 1755 – 1774.

[4] Sandirasegaram N, Liu C. Analysis of polarimetric techniques using high-resolution polarimetry data in an automatic target recognition context [J]. IET Radar, Sonar and Navigation, 2011, 5(2): 163 – 171.

[5] 何松华, 肖怀铁, 孙文峰, 等. 高距离分辨率极化雷达目标匹配识别研究 [J]. 电子学报, 1999, 27(3): 110 – 112.

HE Songhua, XIAO Huaitie, SUN Wenfeng, et al. A study of high range resolution polarization radar target recognition by using matched correlators [J]. Acta Electronica Sinica, 1999, 27(3): 110 – 112. (in Chinese)

[6] 肖怀铁, 郭雷, 付强, 等. 宽带多极化雷达目标模糊匹配识别方法 [J]. 系统工程与电子技术, 2005, 27(5): 770 – 773.

XIAO Huaitie, GUO Lei, FU Qiang, et al. Method of wideband polarization radar target recognition using fuzzy matched filters [J]. Systems Engineering and Electronics, 2005, 27(5): 770 – 773. (in Chinese)

[7] Ma X S, Shen H F, Yang J, et al. Polarimetric-spatial

classification of SAR images based on the fusion of multiple classifiers [J]. IEEE Journal of Selected Topic in Applied Earth Observation and Remote Sensing, 2014, 7(3): 961 – 971.

[8] Paladini R, Martorella M, Berizzi F. Classification of man-made targets via invariant coherency-matrix eigenvector decomposition of polarimetric SAR/ISAR images [J]. IEEE Transaction on Geoscience and Remote Sensing, 2011, 49(8): 3022 – 3034.

[9] Li H J, Lane R Y. Utilization of multiple polarization data for aerospace target identification [J]. IEEE Transaction on Antennas and Propagation, 1995, 43(12): 1436 – 1440.

[10] Cui M S, Prasad S, Mahrooghy M, et al. Decision fusion of textural features derived from polarimetric data for levee assessment [J]. IEEE Journal of Selected Topic in Applied Earth Observation and Remote Sensing, 2012, 5(3): 970 – 976.

[11] Ji S H, Dunson D, Carin L. Multitask compressive sensing[J]. IEEE Transactions on Signal Processing, 2009, 57(1): 92 – 106.

[12] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210 – 227.

[13] Wright J, Ma Y, Mairal J, et al. Sparse representation for computer vision and pattern recognition [J]. Proceedings of the IEEE, 2010, 98(6): 1031 – 1044.

[14] Yuan X T, Liu X B, Yan S C. Visual classification with multitask joint sparse representation [J]. IEEE Transaction on Image Processing, 2012, 21(10): 4349 – 4360.

[15] Zhang H C, Nasrabadi N M, Zhang Y N, et al. Joint dynamic sparse representation for multi-view face recognition[J]. Pattern Recognition, 2012, 45(4): 1290 – 1298.

[16] Liu S Q, Zhan R H, Zhang J, et al. Radar automatic target recognition based on sequential vanishing component analysis [J]. Progress in Electromagnetics Research, 2014, 145: 241 – 250.

[17] 杜兰. 雷达高分辨距离像目标识别方法研究 [D]. 西安: 西安电子科技大学, 2007.

DU Lan. Study on radar HRRP target recognition [D]. Xi'an: Xidian University, 2007. (in Chinese)

[18] Du L, Liu H W, Bao Z, et al. A two-distribution compounded statistical model for radar HRRP target recognition [J]. IEEE Transactions on Signal Processing, 2006, 54(6): 2226 – 2238.

稀疏贝叶斯学习框架下的扩展目标雷达关联成像*

周小利,王宏强,程永强,秦玉亮
(国防科技大学 电子科学与工程学院,湖南 长沙 410073)

摘要:传统的关联成像方法未考虑复杂扩展目标的结构信息,在高分辨成像时的应用受到限制,为此提出一种自适应结构配对稀疏贝叶斯学习方法。该方法在稀疏贝叶斯学习的框架内针对扩展目标建立一种结构配对层次化高斯先验模型,然后采用变分贝叶斯期望-最大化算法交替进行目标重构和参数优化。该方法将某一信号分量的重构与周围信号分量联系起来,并能在迭代过程中自适应地调整表征各信号分量相关性的参数。实验结果表明,该方法针对扩展目标可以有效地进行高分辨成像。

关键词:雷达关联成像;扩展目标;稀疏贝叶斯学习;结构配对;变分贝叶斯期望-最大化

中图分类号:TN957 **文献标志码:**A **文章编号:**1001-2486(2017)03-151-07

Radar coincidence imaging for extended targets in sparse Bayesian learning framework

ZHOU Xiaoli, WANG Hongqiang, CHENG Yongqiang, QIN Yuliang
(College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: Radar coincidence imaging is a high-resolution staring imaging technique without the limitation of relative motion between target and radar. Conventional radar coincidence imaging methods ignore the structure information of complex extended target, which limits its applications in high resolution imaging, thus an adaptive pattern-coupled sparse Bayesian learning algorithm was proposed. To model the extended target, a pattern-coupled hierarchical Gaussian prior model was introduced in sparse Bayesian learning framework, and then the algorithm alternated between steps of target reconstruction and parameter optimization under the variational Bayesian expectation maximization framework. Therefore, the reconstruction of each coefficient involved its immediate neighbors, and the parameter indicating the pattern relevance between the coefficient and its immediate neighbors was updated adaptively during the iterations. Experimental results demonstrate that the proposed algorithm can achieve high resolution imaging effectively for the extended target.

Key words: radar coincidence imaging; extended target; sparse Bayesian learning; pattern-coupled; variational Bayesian expectation maximization

作为一种全天候、全天时、远距离的信息获取手段,雷达成像技术在空间监视、对地观测等领域有着非常重要的应用。现有的高分辨雷达成像系统多采用合成孔径雷达或逆合成孔径雷达成像技术,二者依据距离-多普勒成像原理,属于“运动成像”方式。雷达与目标的相对运动是成像的前提条件,同时也存在复杂运动补偿难度大,在“凝视/近凝视”的非理想观测几何下难以高分辨成像等难题。而雷达关联成像(Radar Coincidence Imaging, RCI)^[1-4]可以与传统的雷达成像系统形成互补。其作为一种新的凝视成像技术,不依赖于雷达与目标的相对运动,具有高分辨、抗截获、抗干扰等优势,在静止/准静止平台凝视成像、灾情监测、海洋监视、高分辨对地观测等应用领域具有广泛的应用前景。

雷达关联成像借鉴经典的光学关联成像思想,通过对发射信号波前的调制,构造在时间和空间上随机分布的二维随机辐射电磁场,以此模拟具有随机涨落的光场分布,然后将目标散射回波与二维随机辐射场进行关联处理,从而实现波束内目标信息的提取与解耦^[1]。

目前,雷达关联成像正受到越来越多的关注和研究。中国科学技术大学^[5]、国防科技大学^[1-3]、西安电子科技大学^[6]、西安交通大学^[4]

* 收稿日期:2016-01-21
基金项目:国家自然科学基金资助项目(61302149, 61302142);高等学校博士学科点专项科研基金博导类资助项目(20124307110013)
作者简介:周小利(1988—),男,湖北随州人,博士研究生,E-mail:zhouxiaoli@nudt.edu.cn;
王宏强(通信作者),男,研究员,博士,博士生导师,E-mail:oliverwhq1970@gmail.com

等单位相继展开相关研究,在关联成像基本原理、超分辨率机理、随机辐射源优化设计、成像算法等方面取得了一系列研究成果。中国科学技术大学利用所研制的原理演示装置验证了关联成像具有超 10 倍天线孔径限制的分辨能力。

由于雷达关联成像模型与压缩感知成像模型存在天然的一致性,目前关联成像算法的研究大多集中在压缩感知/稀疏重构类算法,其前提是目标满足稀疏特性,即目标可由少数占支配地位的局部散射中心近似描述^[7]。在点目标成像中,这类算法的成像效果较好。但是对于较复杂的扩展目标而言,目标散射点较多且呈区域性块聚集特性,空间域的稀疏性相对较差,此时传统的稀疏重构方法所能获得的成像结果常常并不理想^[8]。

对扩展目标进行稀疏成像时,除了要考虑散射点的稀疏先验之外,也要考虑其结构信息,这些信息在传统的稀疏重构算法中通常会被忽略。目前,已经有一些挖掘结构信息的算法,如 Group-BP^[9]、Block-OMP^[10]、Group-LASSO^[11]、Block-CoSaMP^[12]等。这些算法尽管有效,但是需要知道结构信息的先验,例如块的大小和划分,这在实际中通常是无法预知的。稀疏贝叶斯学习(Sparse Bayesian Learning, SBL)由于可以灵活地对目标先验信息进行建模,近来蓬勃兴起。Zhang 等考虑稀疏块元素之间的结构特性提出了块稀疏贝叶斯学习(Block SBL, BSBL)框架^[13]。文献[14]基于群稀疏模型和不同的稀疏先验利用变分贝叶斯参数估计方法提出了变分贝叶斯群稀疏(Variational Bayesian Group-Spare, VBGS)算法。文献[15-16]将“spike-and-slab”先验引入到 SBL 中,基于马尔科夫蒙特卡洛采样(Markov Chain Monte Carlo, MCMC)提出了 CluSS-MCMC^[15]方法,这一先验模型可以在诱导重构信号稀疏性的同时,也提高了信号系数呈块状聚集的可能性。文献[16-17]采用基于图论的先验概率模型来表征元素间的统计相关性。文献[18]同时考虑了整个稀疏信号块的稀疏性和块内元素的稀疏性,提出了一种分层稀疏贝叶斯学习算法(Hierarchical SBL, HiSBL)。文献[19-20]考虑相邻元素之间的统计相关性,提出了一种结构配对(pattern-coupled)的层次化高斯先验模型。在该模型中,各元素的稀疏的先验不仅与自身对应的超参数有关,而且与其相邻元素的超参数有关,这种结构可以有效促进块状聚类特征,挖掘块状先验。

针对传统的关联成像算法对扩展目标的成像

效果不佳的问题,本文提出了一种自适应结构配对稀疏贝叶斯学习算法(Adaptive Pattern-Coupled Sparse Bayesian Learning, APC-SBL)。

1 扩展目标雷达关联成像的贝叶斯模型

1.1 雷达关联成像模型

雷达关联成像的基本原理如图 1 所示。与传统雷达发射相干信号形成平面波前进行成像探测不同,雷达关联成像通过发射特定调制的雷达波形对信号波前进行随机调制,在波束内不同目标处形成具有差异性分布的辐射场激励,确保目标散射回波中蕴含可辨识的空间分布信息,从而为实现波束内超分辨成像提供可能。图 2 为关联成像形成的随机调制波前与传统雷达所形成的平面波前示意图,平面波前在不同时刻形成的波前分布基本一致,多次观测并不能带来信息量的增加;而关联成像的随机调制波前不仅在空间上是随机起伏的,增加了可用于波束内目标分辨的信息,同时在时间上的波前分布也是独立的,从而带来观测信息量的增加,为目标重构和超分辨提供必要条件。

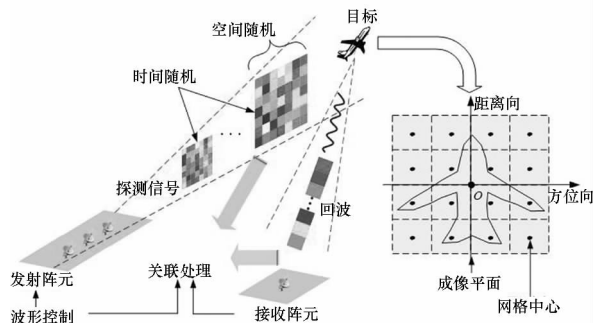
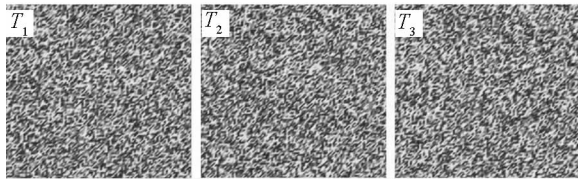


图 1 雷达关联成像原理图

Fig. 1 Basic principle of RCI

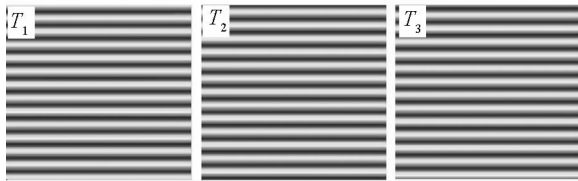
一种比较可行的构造时空二维随机辐射场的方式是通过雷达阵列发射相互独立的随机信号。因此雷达关联成像可以由多发多收(Multiple-Input Multiple-Output, MIMO)雷达系统实现。但是与 MIMO 雷达聚焦于多路径和多观测角不同,雷达关联成像利用发射信号的干涉增强波前的起伏和探测信号的空间差异性,同时 MIMO 雷达利用波形的正交性实现多路径分离,而雷达关联成像不需要进行信号分离。

根据文献[1]给出的模型,采用如图 1 所示的多发单收体制,发射阵元数为 M 。在进行成像处理前,首先将成像平面均匀划分为 K 个网格(也称为成像单元),并假设散射点均位于网格中心上。在雷达关联成像中,各发射阵元发射特定



(a) 不同时刻随机波前

(a) Random wavefront at different times



(b) 不同时刻平面波前

(b) Flat wavefront at different times

图2 探测信号波前示意图

Fig.2 Wavefront of detecting signals

调制的随机波形 $S_{t_m}(t)$, 从而在成像平面上形成时空二维随机辐射场。目标图像可由回波与随机辐射场对应的参考信号进行关联处理得到。根据散射点与随机辐射场的作用规律, 可以将接收回波写为:

$$\begin{cases} y = S \cdot \beta + w \\ \begin{bmatrix} y(t_1) \\ y(t_2) \\ \vdots \\ y(t_N) \end{bmatrix} = \begin{bmatrix} S(t_1, r_1) & \cdots & S(t_1, r_K) \\ S(t_2, r_1) & \cdots & S(t_2, r_K) \\ \vdots & \cdots & \vdots \\ S(t_N, r_1) & \cdots & S(t_N, r_K) \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix} + \begin{bmatrix} w(t_1) \\ w(t_2) \\ \vdots \\ w(t_N) \end{bmatrix} \end{cases} \quad (1)$$

其中: y, β, w 分别表示接收回波、散射系数和高斯噪声矢量; t_n 表示采样时刻。根据散射点均位于网格中心上这一假设, 第 k 个网格中心对应的散射系数为 β_k , $\beta_k = 0$ 表示该网格中心上没有散射点。 S 为各个散射点处的辐射场所组成的参考信号矩阵, 其中第 k 个网格中心 r_k 处的辐射场参考信号为:

$$S(t_n, r_k) = \sum_{m=1}^M S_{t_m} \left(t_n - \frac{|r_k - R_m| + |r_k - R_0|}{c} \right) \quad (2)$$

其中, R_0 和 R_m 分别表示接收阵元和第 m 个发射阵元的位置。

基于目标的稀疏先验, 式(1)表示的成像方程即为典型的稀疏重构模型, 目标的重构可以采用稀疏重构方法来解决。

1.2 扩展目标的贝叶斯模型

基于式(1)中的模型, 雷达关联成像需要利用被噪声污染的数据 y 以及根据成像模型推演的参考信号矩阵 S 来重构散射系数矢量 β 。在对扩展

目标进行关联成像时, 待成像目标的散射点分布是块状稀疏的, 且块结构样式和块大小均未知。为了充分挖掘目标的稀疏先验, 下面在 SBL 框架中, 对目标先验进行建模。

在传统的 SBL 框架中^[21], 一般假设 β 服从高斯-伽马先验分布, 以促进目标的稀疏性。

$$p(\beta | \alpha) = \prod_{k=1}^K CN(\beta_k | 0, \alpha_k^{-1}) \quad (3)$$

$$p(\alpha; a, b) = \prod_{k=1}^K \text{Gamma}(\alpha_k | a, b) \quad (4)$$

其中: $CN(\cdot | \mu, \Sigma)$ 表示均值为 μ 、协方差为 Σ 的复高斯分布, α_k^{-1} 为 β_k 的先验方差, $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_K]^T$ 可以控制 β 的稀疏性。当 α_k 趋于无穷大时, 对应的 β_k 趋于 0。 $\text{Gamma}(\alpha_k | a, b) = \Gamma(a)^{-1} b^a \alpha_k^{a-1} e^{-b\alpha_k}$ 表示 α_k 服从参数为 a 和 b 的伽马分布。

可以看出该模型中 β 中的各个系数之间是相互独立的, 每一个系数由对应的超参数 α_k 来控制, 因此对块状结构没有诱导作用。但实际上, 对于扩展目标而言, 某一散射点的稀疏性可能与周围的散射点有关, 例如当周围的点的散射系数均非 0 时, 该散射点将以较大的概率趋于非 0。因此为了挖掘散射系数的这种相关性, 这里借鉴文献[19]的思想, 针对 β 建立一种结构化贝叶斯模型, 在该模型中, 每个散射系数的先验分布除了与自身对应的超参数有关外, 也与周围散射系数对应的超参数有关, 即

$$p(\beta | \alpha; \rho) = \prod_{k=1}^K CN(\beta_k | 0, \gamma_k^{-1}) \quad (5)$$

其中: $\gamma_k \triangleq \alpha_k + \rho \zeta_k$, $\zeta_k = \sum_{l \in S_k} \alpha_l$, S_k 代表第 k 个网格中心周围所有网格中心元素的集合, ρ 表征该网格与周围各网格的相关性。显然, 当 $\rho = 0$ 时, 该模型退化为传统的 SBL 模型; 当 $\rho \neq 0$ 时, β_k 的稀疏性不仅与 α_k 有关, 也与周围散射系数对应的超参数有关, 这样就使得各个网格对应的散射系数相互关联起来, 从而可以诱导块稀疏。

对于噪声, 通常假设其服从高斯分布, 同时假设其精度(方差的倒数)服从伽马分布(因为伽马分布与高斯分布为共轭分布, 可为后续推导提供便利), 即

$$p(w | \alpha_0) = CN(w | 0, \alpha_0^{-1} I) \quad (6)$$

$$p(\alpha_0; c, d) = \text{Gamma}(\alpha_0 | c, d) \quad (7)$$

其中, α_0^{-1} 为噪声的先验方差, c 和 d 为 α_0 服从的伽马分布的参数。

2 基于 APC-SBL 的扩展目标关联成像算法

基于 1.2 节所建立的模型, 本节将在 VBEM

框架内推导提出的 APC-SBL 算法。

在 VBEM 框架^[22]中,首先定义隐变量(即未观测到的变量)为 $\Omega = \{\beta, \alpha, \alpha_0\}$, 未知的确定性参数为 ρ 。在期望 - 最大化 (Expectation-Maximization, EM) 算法中,首先需要知道隐变量 Ω 的后验分布,即

$$p(\Omega | y; \rho) = \frac{p(\Omega, y; \rho)}{p(y)} \quad (8)$$

其中: $p(\Omega, y; \rho)$ 为联合概率分布,但是由于 $p(y)$ 没有解析表达式,导致传统的 EM 算法不能直接应用。而 VBEM 算法在变分 E 步采用变分贝叶斯推理的方法对隐变量的后验分布进行估计而不是点估计,为此可以通过式(9)估计各个隐变量的近似后验概率密度^[22]。

$$\ln q_k = \langle \ln p(\Omega, y; \rho) \rangle_{i \neq k} + c_0 \quad (9)$$

其中: q_k 表示 Ω 的第 k 个分量 Ω_k 的近似后验概率密度, $\langle \cdot \rangle_{i \neq k}$ 表示相对于 $q_i (i \neq k)$ 的数学期望, c_0 为归一化常数。通过应用 VBEM 算法及以上假设,可以计算出 Ω 中各分量的近似后验分布。利用式(9)可以得到:

$$\begin{aligned} \ln q(\beta) &= \langle \ln p(\Omega, y; \rho) \rangle_{q(\alpha)q(\alpha_0)} + c_0 \\ &= \langle \ln p(y | \beta, \alpha_0; \rho) p(\beta | \alpha; \rho) \rangle_{q(\alpha)q(\alpha_0)} + c_0 \end{aligned} \quad (10)$$

经过推导可得:

$$q(\beta) = CN(\beta | \mu, \Sigma) \quad (11)$$

$$\mu = \langle \alpha_0 \rangle \Sigma S^H y \quad (12)$$

$$\Sigma = (\langle \alpha_0 \rangle S^H S + \langle \Lambda \rangle)^{-1} \quad (13)$$

其中: $\langle \Lambda \rangle = \text{diag}(\langle \gamma_1 \rangle, \langle \gamma_2 \rangle, \dots, \langle \gamma_K \rangle)$ 。

同理,可以推导 α_0 的后验分布,得到:

$$q(\alpha_0) = \text{Gamma}(\alpha_0 | \hat{c}, \hat{d}) \quad (14)$$

$$\hat{c} = c + N - 1 \quad (15)$$

$$\hat{d} = \langle \|\mathbf{S}\beta - \mathbf{y}\|_2^2 \rangle_{q(\beta)} + d \quad (16)$$

式(16)中, $\langle \|\mathbf{S}\beta - \mathbf{y}\|_2^2 \rangle_{q(\beta)} \triangleq \|\mathbf{S}\mu - \mathbf{y}\|_2^2 + \alpha_0^{-1} \sum_{k=1}^K \lambda_k$, 其中 $\lambda_k = 1 - \gamma_k \Sigma_{kk}$, Σ_{kk} 为 Σ 的第 (k, k) 个元素。根据伽马分布的性质, α_0 的后验期望为:

$$\langle \alpha_0 \rangle = \frac{\hat{c}}{\hat{d}} \quad (17)$$

隐变量 α 的后验分布可以通过式(18)计算。

$$\ln q(\alpha) = \langle \ln p(\beta | \alpha; \rho) p(\alpha; a, b) \rangle_{q(\beta)q(\alpha_0)} + c_0 \quad (18)$$

将式(4)和式(5)代入式(18)得:

$$\ln q(\alpha) = \sum_{k=1}^K (a \ln \alpha_k + \ln \gamma_k - \gamma_k \langle |\beta_k|^2 \rangle - b \alpha_k) + c_0 \quad (19)$$

令 $\omega_k \triangleq \langle |\beta_k|^2 \rangle = |\mu|_k^2 + \Sigma_{kk}$ 。从式(19)可以看出与传统 SBL 中各个 α_k 独立更新不同, $\ln q(\alpha)$ 表达式中涉及 α_k 与 $\alpha_l (l \in S_k)$ 的交叉项, 很难求出 α_k 后验分布的准确形式。为此可以考虑在 VBEM 算法中的 M 步中更新 α , 即最大化 E 步中的 Q 函数 $Q(\alpha | \alpha^{\text{OLD}})$ 。 Q 函数表示给出当前估计参数 α^{OLD} 及观测数据时, α 的后验对数概率期望, 这与 $\ln q(\alpha)$ 是一致的。所以

$$Q(\alpha | \alpha^{\text{OLD}}) = \sum_{k=1}^K (a \ln \alpha_k + \ln \gamma_k - \gamma_k \omega_k - b \alpha_k) \quad (20)$$

最大化 Q 函数更新 α 的估计值, 得:

$$\alpha^{\text{NEW}} = \arg \max_{\alpha} Q(\alpha | \alpha^{\text{OLD}}) \quad (21)$$

采用文献[19]介绍的方法, 用简单的次优解来代替最优解进行迭代优化, 可以求得:

$$\alpha_k^{\text{NEW}} = \frac{a-1}{b + \chi_k} \quad (22)$$

其中, $\chi_k = \omega_k + \rho \sum_{l \in S_k} \omega_l$ 。

而对于参数 ρ , 同样可以在 VBEM 中的变分 M 步对其估计进行更新, 即

$$\rho^{\text{NEW}} = \arg \max_{\rho} \langle \ln p(y, \Omega; \rho) \rangle_{q(\beta; \rho^{\text{NEW}})q(\alpha)q(\alpha_0)} \quad (23)$$

经过化简可以得到:

$$\rho^{\text{NEW}} = \arg \max_{\rho} \left(\sum_{k=1}^K \ln \gamma_k - \sum_{k=1}^K \gamma_k \omega_k \right) \quad (24)$$

令 $f(\rho) = \sum_{k=1}^K \ln \gamma_k - \sum_{k=1}^K \gamma_k \omega_k$, 从式(24)可以看出 ρ 的估计是一个非线性问题, 这里采用牛顿迭代法解决该问题, 即

$$\rho^{\text{NEW}} = \rho^{\text{OLD}} - [\nabla^2 f(\rho^{\text{OLD}})]^{-1} [\nabla f(\rho^{\text{OLD}})] \quad (25)$$

其中, $\nabla f(\rho^{\text{OLD}})$ 和 $\nabla^2 f(\rho^{\text{OLD}})$ 分别表示 ρ^{OLD} 处的一阶和二阶导数。 $\nabla f(\rho^{\text{OLD}})$ 和 $\nabla^2 f(\rho^{\text{OLD}})$ 可以化简为:

$$\nabla f(\rho^{\text{OLD}}) = \sum_{k=1}^K \zeta_k \left(\frac{1}{\alpha_k + \rho^{\text{OLD}} \zeta_k} - \omega_k \right) \quad (26)$$

$$\nabla^2 f(\rho^{\text{OLD}}) = - \sum_{k=1}^K \frac{\zeta_k^2}{(\alpha_k + \rho^{\text{OLD}} \zeta_k)^2} \quad (27)$$

在 VBEM 推理过程中, 每步的迭代计算量主要来源于式(12)的矩阵 - 向量乘积和式(13)的矩阵求逆运算, 其计算复杂度分别为 $O(K^2)$ 和 $O(K^3)$, 当网格数 K 较大时, 算法的计算量较大。事实上, 通过网格修剪, 所需处理的网格维数可以不断减小, 从而使得迭代过程中的计算量不断下降。网格修剪的方法是, 当某一网格对应的超参数 α_k^{NEW} 超过设定的阈值 α_{th} 时, 该网格即可被修

剪掉,即

$$H^{i+1} = \{k | \alpha_k^{\text{NEW}} < \alpha_{\text{th}}, k \in H^i\} \quad (28)$$

其中, H^{i+1} 为第 i 次修剪后的网格点的集合。

基于以上分析,利用 APC-SBL 算法重构目标散射系数的流程可以概括如下。

- 1) 初始化: 迭代次数 $i = 0$, $\alpha_0 = 10^2 / \text{VAR}(\mathbf{y})$, $\alpha = N / |\mathbf{S}^H \mathbf{y}|$, 网格点的集合 $H^0 = \{1, 2, \dots, K\}$;
- 2) 更新散射系数: 令 $i = i + 1$, 根据式 (12) 和式 (13) 计算 $\boldsymbol{\mu}$ 和 $\boldsymbol{\Sigma}$;
- 3) 更新参数: 根据式 (22) 计算超参数 α_k^{NEW} ($k \in H^i$), 根据式 (25) 计算 ρ^{NEW} ;
- 4) 修剪网格点: 按照式 (28) 的方法修剪网格点, 同时对 $\boldsymbol{\mu}$ 、 \mathbf{S} 和 α 进行相应的修剪;
- 5) 判断终止条件: 当 $\|\boldsymbol{\mu}^{i+1} - \boldsymbol{\mu}^i\| / \|\boldsymbol{\mu}^i\| \leq \varepsilon$ 或者达到最大迭代次数 I_{max} ;

6) 输出重构结果: 利用当前重构的散射系数 $\boldsymbol{\mu}$ 和网格点集合 H^{i+1} 合成成像结果。

迭代初始值可以影响算法的收敛性能, 参考文献 [8, 19–22] 并结合所提算法和仿真实践对上述流程第 1 步中的初始值进行设置, 大量数值仿真检验了其有效性。根据 VBEM 算法的原理 [22], 在变分 E 步, 以当前隐变量的后验分布 $q^{\text{OLD}}(\boldsymbol{\Omega})$ 与当前参数估计 ρ^{OLD} 作为输入条件, 通过式 (9) 求解新的后验分布 $q^{\text{NEW}}(\boldsymbol{\Omega})$; 在变分 M 步, 通过求解式 (23) 获得新的参数估计 ρ^{NEW} 。APC-SBL 算法在迭代过程中会依次减小 Kullback-Leibler 散度和负对数似然函数的期望, 直至收敛, 同时迭代过程中高阶统计信息的利用也可以减小收敛到局部最小值的可能性。

3 仿真实验

本节将通过数值仿真对所提 APC-SBL 算法在扩展目标关联成像中的性能进行评估。其中评价算法重构性能的指标为: 相对成像误差 (Relative Imaging Error, RIE) $\xi = 20 \lg(\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2 / \|\boldsymbol{\beta}\|_2)$, 其中 $\hat{\boldsymbol{\beta}}$ 和 $\boldsymbol{\beta}$ 分别表示重构的散射系数和真实的散射系数。

雷达工作在 X 波段, 载频为 10 GHz, 发射阵元配置方式为均匀线阵, 阵元数 $M = 8$, 阵元间距为 1 m。各阵元发射随机跳频信号 [23], 信号带宽为 500 MHz。成像平面均匀划分为 40×40 网格, 网格大小为 1 m \times 1 m。APC-SBL 算法的参数 $a = 2$, $c = 1$, $b = d = 10^{-6}$, 网格修剪参数 $\alpha_{\text{th}} = 10^2$, 终止条件参数 $\varepsilon = 10^{-6}$, 最大迭代次数 $I_{\text{max}} = 300$ 。除了所提算法之外, 两种典型的稀疏重构算法——正交匹配追踪 (Orthogonal Matching Pursuit,

OMP) 算法 [24]、平滑 L0 (Smoothed L0-norm, SL0) 算法 [25], 以及三种块稀疏重构算法——Group-BP、BSBL、CluSS-MCMC 算法 [16] 也进行了仿真, 用于对比分析。实验采用的目标模型如图 3 所示, 为相对于简单的点目标, 该目标在距离向和方位向均存在块状结构, 稀疏性相对较差。

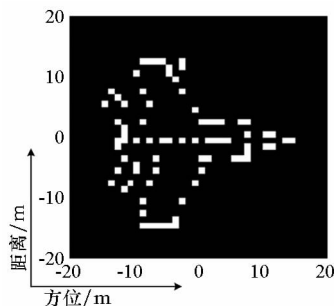


图3 目标场景

Fig. 3 Target scene

实验 1: 算法有效性验证。首先通过仿真实验检验算法对扩展目标的成像效果, 在仿真时加入信噪比为 20 dB 的高斯噪声, 仿真结果如图 4 所示。从图 4 中的成像结果可以看出传统的稀疏重构算法重构性能较差, 存在较多的“虚点”以及不同程度的“散焦”现象。其中 OMP 和 SL0 算法基本无法成像, 这是因为目标的稀疏性较弱, 使 OMP 和基于 L0 范数的 SL0 算法成像性能受限。Group-BP 和 BSBL 算法考虑了目标的块稀疏特性, 但是当块的大小和划分未知时, 很难对块稀疏结构进行精确描述和重构; 因此, 尽管这两种算法所成的图像具有一定的辨识度, 从中可以看出目标轮廓, 但是“虚点”也较多。CluSS-MCMC 算法也考虑了目标的块聚集特征, 并有针对性地进行了建模; 但由于采用的是 Gibbs 采样的方法, 并不能保证收敛到全局最优解, 而且重构性能受噪声影响较大, 因此重构结果也不理想。相比而言, APC-SBL 算法对噪声和块稀疏结构都进行了建模, 可以获得更加清晰、背景更加纯净的目标图像, 几乎不存在明显的“虚点”。

实验 2: 算法性能测试。下面通过蒙特卡洛仿真实验检验本文算法的时效性及对噪声的鲁棒性。所考察的信噪比范围为 10 ~ 40 dB, 对每一个信噪比进行 100 次独立实验, 每次实验记录各个算法的相对成像误差和运行时间, 结果如图 5 所示。仿真时采用的计算机配置为酷睿 i3 处理器, CPU 主频为 3.4 GHz, 内存为 4 GB, 仿真平台为 MATLAB 2012b。

从图 5(a) 中可以看出, 除了 OMP 之外的五

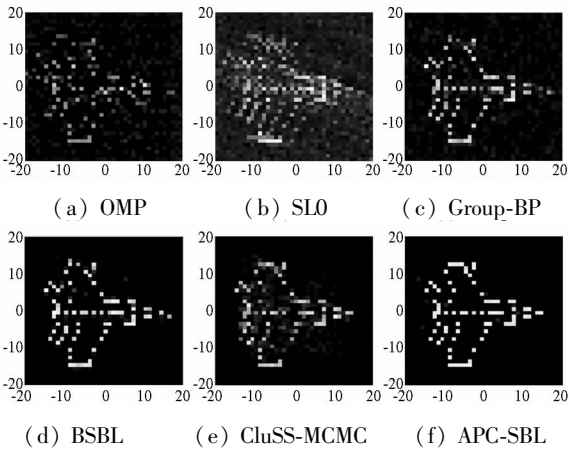
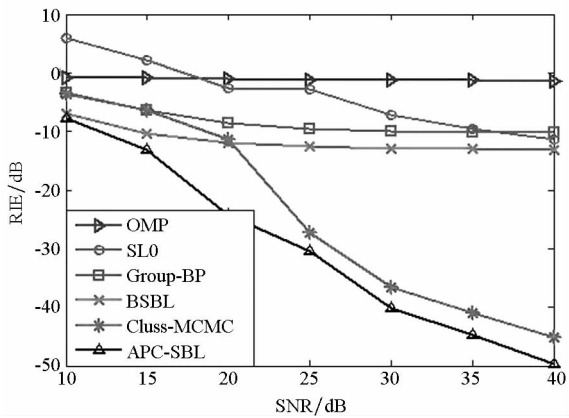


图 4 各种算法的雷达关联成像结果

Fig. 4 RCI results for different algorithms

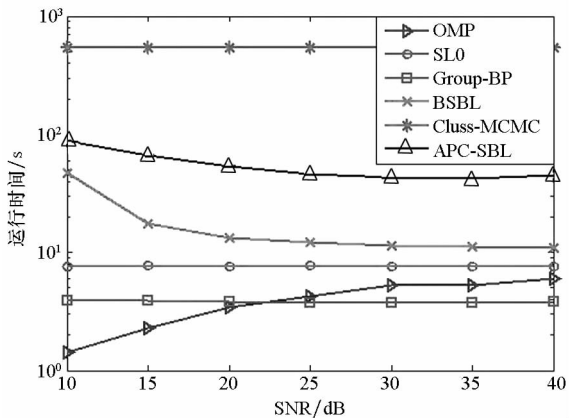
种算法的相对成像误差均随着信噪比的增加而减小,即算法性能与噪声水平密切相关;而噪声对OMP算法的影响较小,其重构误差基本保持在-1 dB左右,此时成像是失败的。同时,APC-SBL算法的相对成像误差随信噪比增加而近似线性递减,说明算法性能对测量噪声比较敏感。由于OMP、SLO算法未考虑目标的块结构先验,重构误差较大。Group-BP、BSBL、CluSS-MCMC和APC-SBL算法考虑了目标的块稀疏结构,重构性能有所改善,特别是在高信噪比时 CluSS-MCMC和APC-SBL两种算法性能优势更加明显。CluSS-MCMC算法尽管在高信噪比时性能较优,但是在低信噪比时性能反而不及BSBL,这是因为算法在更新参数时采用的是Gibbs采样的方法,参数提取精度较低,对噪声比较敏感。相较而言,APC-SBL算法既考虑了目标块稀疏结构,并能根据各次迭代结果自适应地调整各网格之间的相关性参数 ρ ,同时在迭代过程中也对噪声功率进行估计,具有更强的抑制噪声的能力。但是,随着信噪比的降低,APC-SBL算法性能优势越来越不明显,因此如何进一步改善算法在低信噪比下的重构性能是需要进一步研究的内容。

算法的时效性如图5(b)所示。可以看出,OMP和SLO算法运算较快,这两种算法的迭代过程比较简单,计算量小。Group-BP和BSBL算法的计算复杂度较高,计算时间较长,特别是BSBL在低信噪比时的收敛速度变慢,需要的迭代次数增多,运行时间更长。CluSS-MCMC一般需要200次以上的Gibbs采样才能达到稳定采样,收敛速度很慢。APC-SBL算法的时效性虽不及OMP、SLO、Group-BP和BSBL,但是比CluSS-MCMC算法要好,其运行时间与BSBL算法基本处在同一



(a) 信噪比对相对成像误差的影响

(a) RIE vs. SNR



(b) 信噪比对算法运行时间的影响

(b) Runtime vs. SNR

图 5 算法性能测试结果

Fig. 5 Performance test results of proposed algorithm

量级。同时,在低信噪比时,APC-SBL算法达到收敛所需的迭代次数也逐渐增加,算法的时效性变差。

4 结论

作为一种新的凝视高分辨成像体制,雷达关联成像有望突破现有雷达成像系统在非理想观测几何条件下高分辨率成像的瓶颈难题。传统的稀疏重构方法对简单的稀疏点目标成像效果较好。但实际扩展目标的稀疏性往往较差且呈现块聚集特性,此时传统的稀疏重构算法的成像性能会下降。为此本文提出一种基于自适应结构配对稀疏贝叶斯学习的重构算法,针对扩展目标建立一种结构配对层次化高斯先验模型来表征各信号分量间的统计相关性,以诱导块状稀疏结构,然后在VBEM的框架下完成目标重构和参数优化。实验结果表明,与传统的稀疏重构算法相比,所提算法在无须知道目标块状结构的先验信息的情况下,针对复杂的扩展目标依然取得了良好的重构效果。

参考文献 (References)

- [1] Li D Z, Li X, Cheng Y L, et al. Radar coincidence imaging: an instantaneous imaging technique with stochastic signals [J]. IEEE Transactions on Geoscience Remote Sensing, 2014, 52(4): 2261 – 2277.
- [2] Li D Z, Li X, Cheng Y L, et al. Radar coincidence imaging in the presence of target-motion-induced error [J]. Journal of Electronic Imaging, 2014, 23(2): 023014.
- [3] Zhou X L, Wang H Q, Cheng Y Q, et al. Sparse auto-calibration for radar coincidence imaging with gain-phase errors [J]. Sensors, 2015, 15(11): 27611 – 27624.
- [4] Zhu S T, Zhang A X, Xu Z, et al. Radar coincidence imaging with random microwave source [J]. IEEE Antennas and Wireless Propagation Letters, 2015, 14: 1239 – 1242.
- [5] Guo Y Y, He X Z, Wang D J. A novel super-resolution imaging method based on stochastic radiation radar array [J]. Measurement Science and Technology, 2013, 24(7): 074013.
- [6] 邵鹏, 许然, 李浩林, 等. Björck-Schmidt 正交化微波凝视成像方法的研究 [J]. 信号处理, 2014, 30(4): 450 – 456.
SHAO Peng, XU Ran, LI Haolin, et al. The research on Björck-Schmidt orthogonalization for microwave staring imaging [J]. Journal of Signal Processing, 2014, 30(4): 450 – 456. (in Chinese)
- [7] Potter L C, Chiang D M, Carriere R, et al. A GTD-based parametric model for radar scattering [J]. IEEE Transactions on Antennas and Propagation, 1995, 43(10): 1058 – 1067.
- [8] Wang L, Zhao L F, Bi G A, et al. Enhanced ISAR imaging by exploiting the continuity of the target scene [J]. IEEE Transactions on Geoscience Remote Sensing, 2014, 52(9): 5736 – 5750.
- [9] Ewout V D B, Friedlander M P. Probing the pareto frontier for basis pursuit solutions [J]. SIAM Journal on Scientific Computing, 2008, 31(2): 890 – 912.
- [10] Eldar Y C, Kuppinger P, Bölcskei H. Block-sparse signals uncertainty relations and efficient recovery [J]. IEEE Transactions on Signal Processing, 2010, 58(6): 3042 – 3054.
- [11] Eldar Y C, Mishali M. Robust recovery of signals from a structured union of subspaces [J]. IEEE Transactions on Information Theory, 2009, 55(11): 5302 – 5316.
- [12] Baraniuk R G, Cevher V, Duarte M F, et al. Model-based compressive sensing [J]. IEEE Transactions on Information Theory, 2010, 56(4): 1983 – 2001.
- [13] Zhang Z L, Rao B D. Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation [J]. IEEE Transactions on Signal Processing, 2013, 61(8): 2009 – 2015.
- [14] Derin B S, Shinichi N, Do M N. Bayesian group-sparse modeling and variational inference [J]. IEEE Transactions on Signal Processing, 2014, 62(11): 2906 – 2921.
- [15] Yu L, Sun H, Barbot J P, et al. Bayesian compressive sensing for cluster structured sparse signals [J]. Signal Processing, 2012, 92(1): 259 – 269.
- [16] Peleg T, Eldar Y C, Elad M. Exploiting statistical dependencies in sparse representations for signal recovery [J]. IEEE Transactions on Signal Processing, 2012, 60(5): 2286 – 2303.
- [17] Drúmeau A, Herzet C, Daudet L. Boltzmann machine and mean-field approximation for structured sparse decompositions [J]. IEEE Transactions on Signal Processing, 2012, 60(7): 3425 – 3438.
- [18] Wang L, Zhao L F, Bi G R, et al. Hierarchical sparse signal recovery by variational Bayesian inference [J]. IEEE Signal Processing Letters, 2014, 21(1): 110 – 113.
- [19] Fang J, Shen Y N, Li H B, et al. Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals [J]. IEEE Transactions on Signal Processing, 2015, 63(2): 360 – 372.
- [20] Duan H P, Zhang L Z, Fang J, et al. Pattern-coupled sparse Bayesian learning for inverse synthetic aperture radar imaging [J]. IEEE Signal Processing Letters, 2015, 22(11): 1995 – 1999.
- [21] Tipping M E. Sparse bayesian learning and the relevance vector machine [J]. Journal of Machine Learning Research, 2001, 1(3): 211 – 244.
- [21] Tzikas D G, Likas A C, Galatsanos N P. The variational approximation for Bayesian inference [J]. IEEE Signal Processing Magazine, 2008, 25(6): 131 – 146.
- [23] Gogineni S, Nehorai A. Frequency-hopping code design for MIMO radar estimation using sparse modeling [J]. IEEE Transactions on Signal Processing, 2012, 60(6): 3022 – 3035.
- [24] Tropp J A, Gilbert A C. Signal recovery from random measurements via orthogonal matching pursuit [J]. IEEE Transactions on Information Theory, 2007, 53(12): 4655 – 4666.
- [25] Figueiredo M A T, Nowak R D, Wright S J. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems [J]. IEEE Journal of Selected Topics in Signal Processing, 2007, 1(4): 586 – 597.

改进IAHP – CIM模型的雷达组网探测能力评估方法*

崔玉娟, 察 豪

(海军工程大学 电子工程学院, 湖北 武汉 430033)

摘 要:针对雷达组网探测能力的评估问题,采用鱼骨图法对影响雷达组网探测能力的多种因素进行梳理,得到结构层次清晰的指标评价体系。采用改进的区间层次分析—控制区间和记忆模型对该指标体系进行评估,具体方法是:运用区间层次分析法构造区间判断矩阵;利用蛙跳算法求解区间判断矩阵中满足最小一致性的确定性矩阵;利用控制区间和记忆模型来确定各指标的风险概率,实现指标体系的评估。以具体的案例为对象进行仿真实验,结果表明,所提评估方法有效,评估结论相对客观、可信,对雷达组网探测能力评估具有一定参考价值,从而为雷达组网的优化部署奠定良好基础。

关键词:雷达组网;探测能力;鱼骨图;区间层次分析法;蛙跳算法

中图分类号:TN95 **文献标志码:**A **文章编号:**1001 – 2486(2017)03 – 158 – 07

Assessment method for radar network detection capabilities of the improved IAHP-CIM model

CUI Yujuan, CHA Hao

(College of Electronic Engineering, Naval University of Engineering, Wuhan 430033, China)

Abstract: Aiming at evaluating radar network detection capabilities, firstly, a multi-level evaluation index system was established by using the fishbone diagram to analyze various complex factors. Secondly, the IAHP – CIM (interval analytic hierarchy process-controlled intervals and memory) model was proposed to evaluate the index system. Specifically, interval analytic hierarchy process was used to solve the problem of quantified fuzzy index, and the interval judgment matrix was obtained; the Shuffled frog leaping algorithm was applied to optimal interval matrix and the certain number matrix with the minimum consistency ration was obtained; the risk probability of indexes were acquired by the controlled intervals and memory model, and synthetically the index system was evaluated. Finally, simulation results demonstrate the feasibility of the proposed evaluation method, and the relatively objectivity of evaluation conclusion are of great significance to optimize arrangement for radar network.

Key words: radar network; detection capability; fishbone diagram; interval analytic hierarchy process; shuffled frog leaping algorithm

雷达作为防空预警体系的关键装备,在情报信息获取、空中目标引导方面起着至关重要的作用。然而随着科学技术的发展,各种先进武器装备的不断更新,使得雷达所处的战场环境越来越复杂,面临的威胁越来越大,仅仅依靠单部雷达已无法满足现代战争的需求。雷达组网则被认为是应对当前复杂电磁环境的有效手段,可以通过网内不同体制、不同功能、不同频段的多部雷达的相互配合,实现防空预警体系内的情报共享,提高发现目标的精度和速度,增强体系的作战效能。雷达组网探测能力是衡量雷达组网作战效能的重要因素,因此,采用有效的方法来评估雷达组网的探测能力,对评估雷达组网作战效能有着重要的

意义,同时还可以为雷达组网的优化部署提供一个重要的指标参数^[1-2]。

评估雷达组网探测能力是一个涉及多因素、多学科的复杂问题。对此,研究人员做了大量的研究工作,基本思想都是以实现特定目标为前提,按照与之相关的原则,构建指标评价体系^[3-4],采用合理的方法对该指标体系进行评估。文献[5]从雷达组网防空作战效能的定义出发,运用层次分析法(Analytic Hierarchy Process, AHP)分析各要素,并建立各要素的支配关系,进而构建递阶层次的评估指标体系;文献[6]从雷达组网作战的战术、技术角度分析,并融合雷达专家经验,建立评价指标体系。代表性的评估方法有:李莎澜

* 收稿日期:2016 – 02 – 03

基金项目:国家自然科学基金资助项目(41405009);国家重点实验室基金资助项目(K201510)

作者简介:崔玉娟(1984—),女,江苏睢宁人,博士研究生,E-mail:xiaoxiao926878@126.com;

察豪(通信作者),男,教授,博士,博士生导师,E-mail:310938289@qq.com

等^[5]应用模糊层次分析法来评估雷达组网作战效能,通过层次分析法来确定评价体系中的各种指标权重,但是对于体系中部分存在灰色性的指标参数的权重确定,该方法就略显不足。针对这些存在灰色性指标参数的权重确定,胡宗辉等^[6]则采用灰色层次分析法来解决,通过定位区间灰数求解白化矩阵的方法来构造判断矩阵,并对构造的灰色判断矩阵作合理的变换,使其不需要满足一致性检验的要求,并通过实验验证了该方法的有效性。但是选取灰色判断矩阵的定位系数、确定专家权重系数需要先验性,具有较大的主观性。

对此,本文针对评估雷达组网探测能力的问题,采用鱼骨图法的思想对该问题的影响因素进行梳理,将问题剖析分解反映在鱼骨的构架上,从而建立一个结构清晰、层次关系分明的指标评价体系。

1 雷达组网探测能力评价指标体系的建立

鱼骨图^[7]即为鱼的骨架图,鱼骨图法则是通过“诊断”复杂问题,从不同部位进行“号脉”,将复杂问题分解为若干子问题,再根据需要进行“号脉”子问题进行细分,如此反复直至达到最终的目的。图1是鱼骨图的基本构架,其中图1(a)是鱼骨图的基本构架,图1(b)和图1(c)分别是鱼骨图的局部以及对应的名称,图1中标号的含义如下:①特性;②主骨(用粗线和箭头绘制而成);③要因;④大骨(与主骨呈一定角度,比如60°夹角)⑤中骨(与主骨平行);⑥小骨(与中骨呈一定角度,如60°夹角);⑦重要因素。

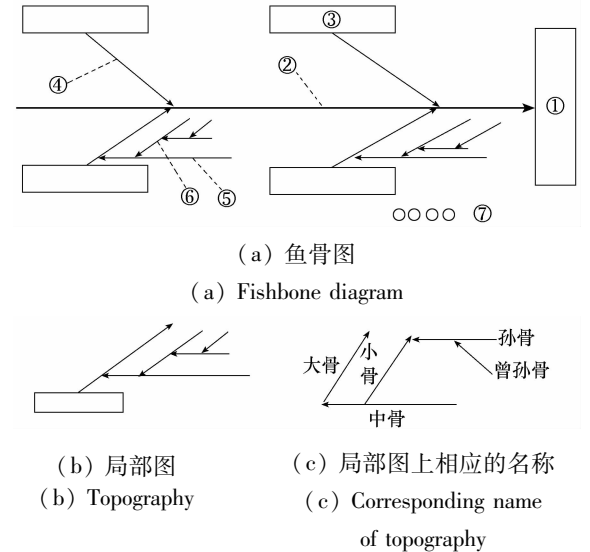


图1 鱼骨图的基本构架

Fig.1 Basic framework of fishbone diagram

雷达组网探测能力按照作战需要具有多样性的特点,如对低空/超低空目标的探测能力、对中空高空目标的探测能力、对隐身目标的探测能力等,而这些细分的探测能力又需要多种指标来评估。因此,如何将这些因素按照一定的原则进行分类整理并进行定量处理,是综合评定雷达组网探测能力的关键。

在目的性、独立性、敏感性、有限性和可实施性的原则下,按照鱼骨图的思想,第一步确定要解决的问题——评估雷达组网探测能力。第二步对确定的问题进行“诊断”处理,将其分解成以下几个子问题:中空高空目标探测能力、低空/超低空目标探测能力、隐身目标探测能力、复杂环境下目标探测能力。第三步对每个子问题进行“号脉”处理,进行进一步的细分,直至分解成满足要求的指标为止。

在第二步中,对中空高空目标探测能力“号脉”时,从雷达探测区域与责任区域之间的关系、雷达之间探测区域的关系角度出发,细化为中空高空空域覆盖率、重叠率以及探测增值率,分别反映了雷达组网覆盖空域的连续性、严密性以及超出责任区的探测能力,具体如图2所示。

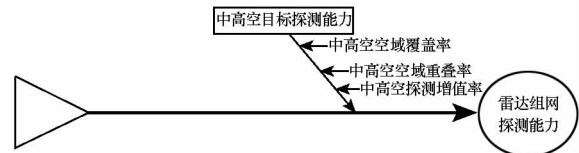


图2 中空高空目标探测能力

Fig.2 Detection capability for medium-altitude targets

在“号脉”低空/超低空目标探测能力时,除与中空高空目标探测能力相同的出发点,提出低空/超低空空域重叠率、覆盖率。还需注意到网内雷达的配置、地面杂波的影响,引入网内雷达的杂波可见度的均值来表示单部雷达低空/超低空探测能力,从而细分为雷达类型因子和单部雷达低空/超低空探测能力,如图3所示。

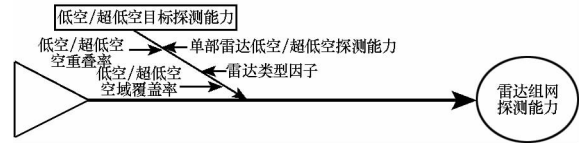


图3 低空/超低空目标探测能力

Fig.3 Detection capability for low altitude/extreme low altitude targets

对隐身目标探测能力进行“号脉”时,从雷达组网的定义出发,网内的雷达是不同体制、不同频段、不同程式、不同极化方式的,从上述方面提取

出空域、频域和极化域等反隐身能力指标;同时,通过通信手段将雷达链接成网,需要考虑信息融合的作用以及整个网内雷达反隐身能力的均值,分别提取出信息融合反隐身能力和单部雷达反隐身能力等指标,见图 4。

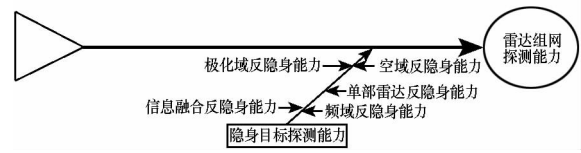


图 4 隐身目标探测能力

Fig. 4 Detection capability for stealth targets

在“号脉”复杂环境下目标探测能力时,复杂环境包括大气、地形、干扰等影响因素,这时需要考虑网内的单部雷达抗复杂环境能力。同时网内雷达的多样性造就了空域、频域上的重叠,极化类型、信号类型的多样性,从而可以相应地提出指标:空域重叠率、频域重叠率、极化类型因子、信号类型因子,如图 5 所示。

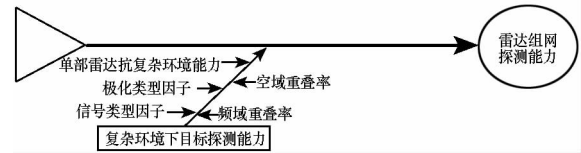


图 5 复杂环境下目标探测能力

Fig. 5 Detection capability for targets in complex environments

通过上述分析,可建立雷达组网探测能力影响因素的鱼骨图,如图 6 所示。

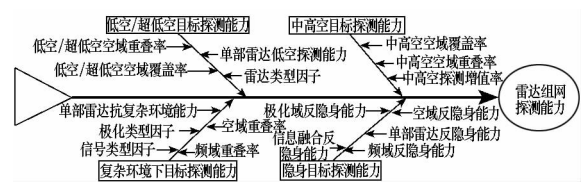


图 6 雷达组网探测能力影响因素的鱼骨图

Fig. 6 Fishbone diagram of influencing factors of radar network detection capabilities

2 雷达组网探测能力评价方法

针对上节建立的雷达组网探测能力的指标评估体系,采用改进的 IAHP - CIM 模型对该指标体系进行评估,其中针对指标体系中的部分指标难量化,容易产生判断不一致、可信度低等问题,将区间数引入到 AHP 中,用区间数来表示指标的权重;但是区间数给出的是模糊指导,并非确定的、具体的,所以利用 SFLA 寻优,得到满足条件的指标的权重点值;最后建立起雷达组网探测能力与风险之间的映射关系,通过计算风险评估得到组

网探测能力的评估结果。图 7 为建立 IAHP - CIM 模型的分析过程。

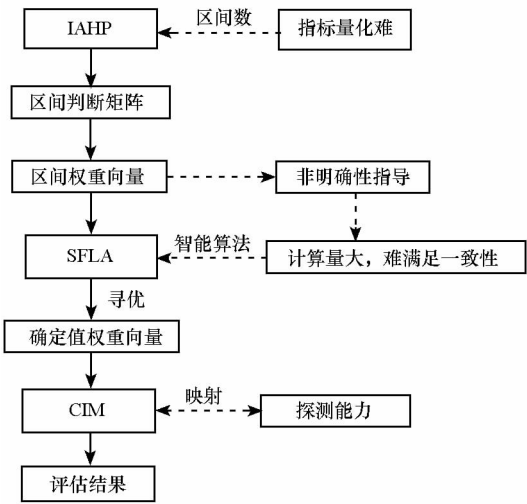


图 7 IAHP - CIM 的分析过程

Fig. 7 Analysis process of IAHP - CIM

2.1 区间层次分析法

IAHP 可看成是 AHP 的变形,主要区别是 IAHP 中涉及的判断矩阵以及其最大特征值对应的特征向量都是区间数,而 AHP 中涉及上述内容的数据都是一个确定的数值。因此 IAHP 的本质就是一个融合了区间数^[8]的 AHP。

雷达组网探测能力的评价体系中的指标带有模糊性和不确定性,多数指标是定性指标,比如反隐身能力,其评价结果可能就是“优”“良”“中”“差”等。如果用 AHP 中确定数值很难描述,但是 IAHP 就能很好地采用区间数来解决,从而提高了算法的客观性和实用性。

虽然 IAHP 法得到的结果数据也为区间数,增强了指标评价的客观性,降低了主观意向对指标评价的影响,但是对于明确性的指导意义不大,所以该方法更多地被用于多目标决策与排序,而非评价目标。对于该问题,不少学者通过区间数特征根法求得区间权重,再取区间两端点值的平均数作为点值权重,得到确定性权重,从而有效地解决上述问题,但此时得到的权重值并不一定是满足一致性条件下的最优解。

2.2 SFLA 确权算法

针对 IAHP 法中通过区间数特征根法求区间权重,然后取均值得到的确定性权重不能保证是满足一致性最优解的问题,因此引入 SFLA^[9] 寻优:首先,将最小的一致性比例作为优化约束条件;其次,对区间判断矩阵进行 SFLA 寻优,得到满足最小一致性条件的确定数判断矩阵;最后,采

用和积法求出该确定数判断矩阵的点值权重向量。通过该方法不仅可以得到最优的确定性权重,还可以较好地解决一致性问题,同时寻优的结果也可以直接评估雷达组网探测能力评价指标体系。

介绍上述提及一致性比例的定义和相关的结论,以方便理解 SFLA 算法的应用。

定义 1 $C.R.(A) = \frac{\lambda_{\max}(A) - n}{(n-1)R.I}$

式中, $C.R.(A)$ 表示矩阵 A 的一致性比例, $\lambda_{\max}(A)$ 表示矩阵 A 的主特征值, $R.I$ 表示随机一致性指标, n 表示矩阵 A 的行数或者列数。

定义 2 若 $C.R.(A) \leq 0.1$ 时,则称矩阵 A 具有满意的一致性。

结论:如果区间数判断矩阵合理地限定,则具有最小一致性比例的判断矩阵是存在并且是唯一的^[10]。

上述结论指出了利用 SFLA 寻优确定权重是有意义的。

SFLA 确权依据为:一般判断效果越好要求其一致性程度越高,从而要求决策者的逻辑判断一致性程度越好,一致性比例数值越小。当某矩阵满足定义 2 时,其求出的权重才可以作为评价指标的权重,矩阵的一致性比例数值越小,其计算出来的评价结论的可靠性就越高。

SFLA 确权思路为:区间数判断矩阵按均匀分布概率随机生成 N 个确定数判断矩阵 $A^k = (a_{ij})_{n \times n}^k, k=1,2,\dots,N$, 满足 $a_{ij} \in [a_{ij}^-, a_{ij}^+], a_{ji} = 1/a_{ij}, a_{ii} = 1$ 。记 $C.R.(A^g) = \min_{k=1,\dots,N} \{C.R.(A^k)\}$, 建立权重数学模型,即:

$$\begin{cases} \min \{C.R.(A^k)\} \\ \text{s. t. } A^k = (a_{ij})_{m \times n}^k \\ a_{ij}^k \in a_{ij} \\ A^k \omega^k = \lambda_{\max}(A^k) \omega^k \\ C.R.(A^k) \leq 0.1 \end{cases} \quad (1)$$

式中, ω^k 为矩阵 A^k 的最大特征值 λ_{\max} 对应的特征向量。

为了在 SFLA 中应用方便,引入一个较大常数 G ,使得 $G - C.R.(A^k) > 0$,这时目标函数从 $\min_{k=1,\dots,N} \{C.R.(A^k)\}$ 转化为 $\max_{k=1,\dots,N} \{G - C.R.(A^k)\}$ 。

2.3 CIM 模型

CIM 模型是 Chapman 和 Cooper 研究概率分布有效叠加的基础上提出的。它可简单直观地展现风险因素的量化过程,也可对风险指标的综合

叠加的叠加误差有效控制^[11-12],该模型包括两种响应模型——串联、并联响应模型。这里仅介绍并联响应模型,如图 8 所示。对于具有多指标影响因子的活动,将计算出第一、第二指标的概率组合的结果与第三个指标做概率组合运算,依次类推,直至最后一个指标。上述过程的概率组合表达式为:

$$P(X_a = x_a) = \sum_{i=1}^m P(X_1 = x_a, X_2 = x_i) + \sum_{i=1}^m P(X_1 = x_{i-1}, X_2 = x_a) \quad (a = 1, 2, \dots, m) \quad (2)$$

式中: X_1, X_2 为 2 个风险因素; x_a 为概率区间的组中值; m 为分组数。

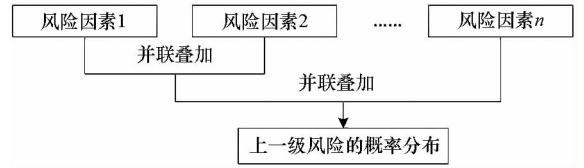


图 8 并联响应模型

Fig. 8 Parallel response model

该模型中的各个变量是相互独立的,变量之间的相关性通过主观概率数值体现,故不再考虑计算方面。对雷达组网探测能力的评估可以看成是对雷达组网投入使用风险的评估,若探测能力强,则其对应的投入的风险小。因此,可以用 CIM 模型来评估雷达组网探测能力。

2.4 雷达组网探测能力评价步骤

针对图 6 中采用鱼骨图法建立的雷达组网探测能力指标评估体系,采用改进的 IAHP - CIM 模型进行评估,其流程如图 9 所示。具体方法如下:

步骤 1:确定“要因”指标层的区间判断矩阵,并运用 SFLA 求出一致性比例最小的确定数判断矩阵。由 1~9 标度法给出评判指标的区间数,从而建立区间数判断矩阵。设某层次有 n 个指标,则该层的区间数判断矩阵表示为:

$$\bar{A} = \begin{bmatrix} \overline{a_{11}} & \cdots & \overline{a_{1j}} & \cdots & \overline{a_{1n}} \\ \vdots & & \vdots & & \vdots \\ \overline{a_{i1}} & \cdots & \overline{a_{ij}} & \cdots & \overline{a_{in}} \\ \vdots & & \vdots & & \vdots \\ \overline{a_{n1}} & \cdots & \overline{a_{nj}} & \cdots & \overline{a_{nn}} \end{bmatrix}$$

利用 SFLA 对 \bar{A} 寻优,即寻找具有最小一致性比例的确定数矩阵 A^g 。假设在 D 维解空间中,第 i 个青蛙的位置代表第 i 个可行解,表示为 $P_i =$

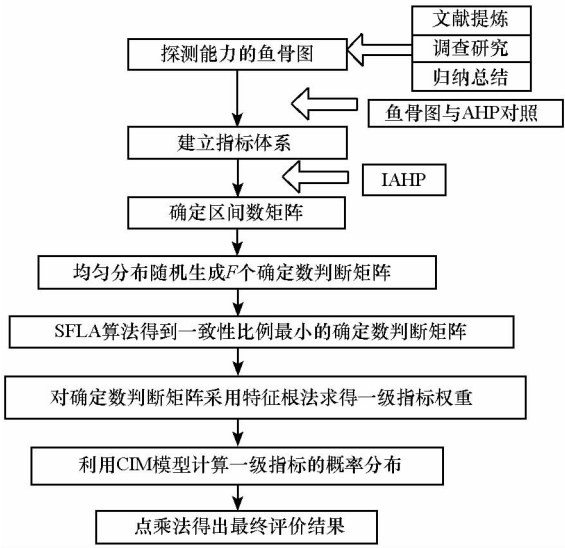


图9 指标体系评价流程

Fig.9 Flow chart of index evaluation system

$(p_{i1}, p_{i2}, \dots, p_{id})$ 。整个蛙群中位置最好的青蛙记为 P_g , 将青蛙等分成若干组, 记 P_b 和 P_w 分别为每次进化时组内位置最好的青蛙和位置最差的青蛙, 则每次进化时, 仅仅对最差的青蛙个体 P_w 实施更新策略, 更新方式为:

$$\begin{cases} D_i = \alpha \cdot (P_b - P_w), & (|D_{\min}| \leq |D_i| \leq |D_{\max}|) \\ P_w^{\text{new}} = P_w + D_i \end{cases} \quad (3)$$

式中: $D_i (i=1, 2, \dots, P)$ 为青蛙个体的更新步长; D_{\min} 为最小更新步长; D_{\max} 为最大更新步长; α 为均匀分布在 $[0, 1]$ 之间的随机数。

如果更新后的 P_w^{new} 的值优于 P_w , 则用 P_w^{new} 替代 P_w , 否则执行如下更新策略, 即:

$$\begin{cases} D_i = \alpha \cdot (P_g - P_w), & (|D_{\min}| \leq |D_i| \leq |D_{\max}|) \\ P_w^{\text{new}} = P_w + D_i \end{cases} \quad (4)$$

如果 P_w^{new} 的值仍然没有改进, 则随机生成一只青蛙 P_r 来代替 P_w 。

有几点需要说明: ①“位置最好”中的位置是指青蛙的适应度, 即青蛙的相应的适应度函数 $\max \{G - C.R.(A^k), k=1, 2, \dots, N\}$ 值; ②经历更新策略得到的 P_w^{new} 的某个分量超过区域范围, 则以靠近其的上限值或下限值代替这个分量, 以保证解的有效性。

步骤 2: 利用特征根法计算 A^g 最大特征值对应的特征向量, 即“要因”指标的权重。

步骤 3: 计算图 6 中的“要因”指标层各指标的概率分布。首先建立雷达组网探测能力与风险之间的映射关系, 接着给出“中骨”指标层中的各

指标的风险概率分布, 最后对“中骨”指标层各指标利用式(2)并联叠加计算, 最终得出“要因”中各指标的最终概率分布。

步骤 4: 利用步骤 2 和步骤 3 的结果求得雷达组网探测能力的评价结果。

3 实例验证

在四川盆地东部, 长江北岸, 地形复杂, 现拟部署 10 部雷达构成组网, 且满足以下条件: 雷达体制数为 5, 工作频段数为 10, 极化类型数为 4, 调制方式数量为 7, 采用圆弧形配置, 雷达组网重叠系数为 2。上述雷达组网效费比略高。首先如图 6 所示的指标体系, 根据 1~9 标度法给出“要因”中各指标相互间的重要程度, 表示为如下的区间数判断矩阵:

$$\bar{A} = \begin{bmatrix} [1, 1] & [2, 5] & [2, 4] & [1, 3] \\ [1/5, 1/2] & [1, 1] & [1, 3] & [1, 2] \\ [1/4, 1/2] & [1/3, 1] & [1, 1] & [1/2, 1] \\ [1/3, 1] & [1/2, 1] & [1, 2] & [1, 1] \end{bmatrix}$$

其次, 利用 SFLA 对 \bar{A} 进行寻优。利用 MATLAB 编程 SFLA, 其参数设置如下: 群内青蛙为 50 个, 分成 5 组, 每组 10 个青蛙, 组内进化次数为 10, 最大迭代次数为 50, $D_{\max} = 0.25$ 。最终得到寻优结果 A^g 为:

$$A^g = \begin{bmatrix} 1 & 3 & 4 & 5 \\ 1/3 & 1 & 1 & 2 \\ 1/4 & 1 & 1 & 1 \\ 1/5 & 1/2 & 1 & 1 \end{bmatrix}$$

接着, 根据定义 1 计算出 $C.R.(A^g) = 0.0137$ 小于 0.1, 满足定义 2, 一致性检验通过。再计算矩阵 A^g 的最大特征值对应的特征向量, 即得到“要因”指标层中各指标的权重向量为:

$$\omega^T = (0.458\ 3, 0.283\ 6, 0.143\ 9, 0.114\ 2)$$

然后, 建立雷达组网探测能力与风险之间的映射关系。建立雷达组网探测能力评价集 $V = \{90 \sim 100, 80 \sim 89, 70 \sim 79, 60 \sim 69, 60 \text{ 分以下}\}$, 其相应的风险评价集为 $V' = \{\text{风险低, 风险较低, 风险适中, 风险较高, 风险高}\}$ 。根据鱼骨图分析每个指标, 结合专家投票指导意见, 得到它们的概率分布, 如表 1 所示。根据式(2)对“中骨”指标层中指标并联叠加计算, 最终得到“要因”指标层中指标在评价集的概率分布。这里只给出 B_1 的概率分布计算过程, 如表 2 所示。其他指标类似求出, 结果如表 3 所示。

表 1 二级指标的概率分布表

Tab. 1 Probability distribution of the second-class index

指标	探测等级				
	90 ~ 100	80 ~ 89	70 ~ 79	60 ~ 69	0 ~ 59
C ₁₁	5/20	7/20	6/20	2/20	0/20
C ₁₂	9/20	7/20	4/20	0/20	0/20
C ₁₃	3/20	11/20	6/20	0/20	0/20
C ₂₁	0/20	2/20	9/20	6/20	3/20
C ₂₂	2/20	5/20	8/20	4/20	1/20
C ₂₃	0/20	3/20	9/20	5/20	3/20
C ₂₄	0/20	3/20	8/20	5/20	4/20
C ₃₁	5/20	6/20	7/20	2/20	0/20
C ₃₂	8/20	7/20	4/20	1/20	0/20
C ₃₃	5/20	6/20	8/20	1/20	0/20
C ₃₄	2/20	5/20	9/20	4/20	0/20
C ₃₅	3/20	6/20	8/20	3/20	0/20
C ₄₁	2/20	5/20	7/20	5/20	1/20
C ₄₂	4/20	5/20	6/20	4/20	1/20
C ₄₃	2/20	3/20	5/20	7/20	3/20
C ₄₄	0/20	2/20	7/20	9/20	2/20
C ₄₅	1/20	5/20	7/20	6/20	1/20

根据表 3 的结果,记

$$B = \begin{bmatrix} 0.016\ 9 & 0.319\ 1 & 0.564\ 0 & 0.100\ 0 & 0.000\ 0 \\ 0.000\ 0 & 0.000\ 8 & 0.147\ 7 & 0.434\ 9 & 0.416\ 6 \\ 0.000\ 4 & 0.035\ 4 & 0.516\ 6 & 0.447\ 6 & 0.000\ 0 \\ 0.000\ 0 & 0.001\ 2 & 0.075\ 6 & 0.579\ 1 & 0.344\ 1 \end{bmatrix}$$

最后,将求出来“要因”指标的权重与其对应的概率分布乘积作和,得到雷达组网探测能力对应的评价等级的概率分布 P 为:

$$P = \omega^T \cdot B$$
$$= (0.007\ 8, 0.151\ 7, 0.383\ 3, 0.299\ 7, 0.157\ 4)$$

上述计算结果表明,组雷达组网的中高空目标探测能力较强(比重为 0.458 3),抗复杂环境能力较弱(比重为 0.114 2)。而最终的评价结果概率分布显示该雷达组网的探测能力等级为 [70,80) 的可能性最大,概率为 38.33%。可根据探测任务和目的来决定雷达组网是否可以投入使用,若不满足使用条件,可适当调整网内资源(如选择抗隐身能力强的雷达)。因此,所提模型不仅能具体分析出每一个子问题,综合给出组网探测能力等级分布概率,而且还对网内资源的配置和调整给出指导。

表 2 B_1 的概率分布计算过程

Tab. 2 Computational process of probability distribution of B_1

探测等级	C ₁₁ 与 C ₁₂ 组合 概率分布	最终概率分布
90 ~ 100	(5/20) × (9/20) = 9/80	(9/80) × (3/20) = 0.016 9
80 ~ 89	(7/20) × (9/20 + 7/20) + (7/20) × (5/20) = 0.367 5	0.367 5 × (11/20 + 3/20) + (11/20) × (9/80) = 0.319 1
70 ~ 79	(6/20) × (9/20 + 7/20 + 4/20) + (4/20) × (7/20 + 5/20) = 0.42	(6/20) × (9/80 + 0.367 5) + 0.42 = 0.564 0
60 ~ 69	(2/20) × 1 = 0.100 0	0.1
0 ~ 59	0	0

表 3 一级指标的概率分布表

Tab. 3 Probability distribution of the first-class index

指标	探测等级				
	90 ~ 100	80 ~ 89	70 ~ 79	60 ~ 69	0 ~ 59
B_1	0.016 9	0.319 1	0.564 0	0.100 0	0.000 0
B_2	0.000 0	0.000 8	0.147 7	0.434 9	0.416 6
B_3	0.000 4	0.035 4	0.516 6	0.447 6	0.000 0
B_4	0.000 0	0.001 2	0.075 6	0.579 1	0.344 1

4 结论

评估雷达组网探测能力是雷达组网优化部署的前提,而评估探测能力的关键在于探测能力指标体系合理有效的建立和评估方法的选用。本文采用鱼骨图法构建一套合理的雷达组网探测能力指标评估体系,并采用 IAHP – CIM 模型对该结构体系进行有效的评估,较好地解决了指标体系中一些定性指标的评估问题,评估结果客观、真实可信,对雷达组网探测能力的评估以及雷达的优化部署具有一定的参考价值。

参考文献(References)

[1] 崔玉娟, 察豪, 田斌. 改进的混合蛙跳算法在雷达网部署中的应用[J]. 海军工程大学学报, 2015, 27(1): 108 – 112.
CUI Yujuan, CHA Hao, TIAN Bin. Improved shuffled frog leaping algorithm for radar network deployment[J]. Journal of Naval University of Engineering, 2015, 27(1): 108 – 112. (in Chinese)

[2] Cui Y J, Cha H, Tian B. Cultural shuffled frog leaping algorithm and its applications for radar network[J]. Applied

- Mechanics & Materials, 2014, 624: 516 – 519.
- [3] 朱丽莉, 王朝晖. 基于模糊层次分析法的雷达组网作战效能评估[J]. 战术导弹技术, 2003(2): 61 – 65.
ZHU Lili, WANG Zhaochi. Evaluation model of fighting effectiveness of radar netting based on FAHP[J]. Tactical Missile Technology, 2003(2): 61 – 65. (in Chinese)
- [4] 周琳, 徐进, 马艳琴, 等. 雷达组网探测系统综合效能评估方法研究[J]. 电子工程师, 2007, 33(9): 10 – 13.
ZHOU Lin, XU Jin, MA Yanqin, et al. Research on evaluation method of radar-netted detection system's integrated effectiveness[J]. Electronic Engineer, 2007, 33(9): 10 – 13. (in Chinese)
- [5] 李莎澜, 刘清国, 魏文斌, 等. 应用模糊层次分析法评估雷达组网作战效能[J]. 湖北工业大学学报, 2010, 22(1): 91 – 93.
LI Shalan, LIU Qingguo, WEI Wenbin, et al. Application of FAHP in the evaluation model of fighting effectiveness of radar netting[J]. Journal of Hubei University of Technology, 2010, 22(1): 91 – 93. (in Chinese)
- [6] 胡宗辉, 钱建刚, 李月岗, 等. 基于改进的 GAHP 确定雷达组网作战效能的指标权重[J]. 兵工自动化, 2009, 28(4): 26 – 28.
HU Zonghui, QIAN Jiangang, LI Yuegang, et al. Index weight determination of operational effectiveness in radar netting based on improved GAHP[J]. Ordnance Industry Automation, 2009, 28(4): 26 – 28. (in Chinese)
- [7] 朱天宇, 孙明. 基于鱼骨图及主成分分析社区公共安全承载力与规划管理对策[J]. 灾害学, 2015, 30(2): 215 – 219.
ZHU Tianyu, SUN Ming. Capacity and planning management measure of community public safety based on fishbone diagram and principal component analysis[J]. Journal of Catastrophology, 2015, 30(2): 215 – 219. (in Chinese)
- [8] 肖峻, 王成山, 罗凤章. 区间层次分析法的权重求解方法初探[J]. 系统工程与电子技术, 2004, 26(11): 1597 – 1600.
XIAO Jun, WANG Chengshan, LUO Fengzhang. Exploration on the methods of weight calculation in the interval-based AHP[J]. Systems Engineering and Electronics, 2004, 26(11): 1597 – 1600. (in Chinese)
- [9] Cui Y J, Cha H, Shen H. Shuffled frog leaping applied to optimal deployment of radar network[J]. Applied Mechanics & Materials, 2014, 624: 512 – 515.
- [10] Finan J, Hurley W. Analytic hierarchy process: does adjusting a pairwise comparison matrix to improve the consistency ratio help? [J]. Computers & Operations Research, 1997, 24(8): 749 – 755.
- [11] 陈阳. 高速公路 BOT 融资项目全过程风险 CIM 模型评价研究[D]. 长沙: 长沙理工大学, 2012.
CHEN Yang. The highway BOT financing project risk assessment based on CIM model[D]. Changsha: Changsha University of Science and Technology, 2012. (in Chinese)
- [12] 高祺勋, 权聪娜, 李博, 等. 灰色模糊 CIM 模型的电力项目融资风险评判构架[J]. 工业工程, 2010, 13(4): 96 – 99.
GAO Qixun, QUAN Congna, LI Bo, et al. Financing risk assessment of power plant construction by using grey fuzzy CIM model[J]. Industrial Engineering, 2010, 13(4): 96 – 99. (in Chinese)

磁悬浮反作用飞轮高精度力矩控制*

冯 健,刘 昆,冯昱澍
(国防科技大学 航天科学与工程学院,湖南 长沙 410073)

摘要:为提高磁悬浮反作用飞轮输出力矩精度,针对传统控制方法下永磁无刷直流电机非理想反电势和换相引起的转矩脉动,分别提出补偿控制策略。非换相期间,根据转速和位置信息,估计实时反电势来获取参考电流,通过设计的力矩控制器直接计算出调制占空比以补偿非理想反电势引起的力矩脉动;分析全转速范围内换相期间转矩波动的特点,分别提出低速区非换相相调制和中高速区间关断相调制的换相转矩脉动抑制策略,并给出换相时间的计算方法。实验表明所提控制方法显著提高了飞轮的输出力矩精度,从而验证了方法的正确性和有效性。

关键词:非理想反电势;换相;转矩脉动;补偿方法

中图分类号:V42 **文献标志码:**A **文章编号:**1001-2486(2017)03-165-07

High-precision torque control for magnetically
suspended reaction flywheel

FENG Jian, LIU Kun, FENG Yushu

(College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: In order to improve the torque-output precision of the magnetically suspended reaction flywheel, compensation methods were proposed respectively to attenuate the undesirable torque ripple caused by nonideal back electromotive force waveform and commutation in classical control of brushless direct current motor. The real-time back electromotive force was estimated according to the rotor position and speed in order to obtain the reference current, the pulse width modulation duty cycle was calculated in the torque controller to compensate the torque ripple caused by the nonideal back electromotive force. The theoretical derivation was analyzed, methods of modulation of the non-commutation phase during low-speed range and modulation of the switching-out phase during high-speed range were presented on the basis of the characteristics respectively. Also, a calculation method of the commutation time was given. The experimental results show that the proposed methods can achieve an effective compensation.

Key words: nonideal back electromotive force; commutation; torque ripple; compensation method

反作用飞轮是航天器姿态控制系统的关键执行机构^[1]。其主要功能是复现航天器姿态控制系统给出的力矩指令,为航天器提供反作用力矩。随着对航天器姿态控制精度要求越来越高,飞轮被要求具有较高的输出力矩精度。磁悬浮反作用飞轮输出力矩精度主要由电机控制精度决定,抑制电磁力矩脉动是获得高精度电磁力矩的重点。无铁芯无齿槽无刷直流电机(BrushLess Direct Current Motor, BLDCM)能量密度高、控制简单且有效地减小了铁芯损耗和磁阻力矩波动,在航天用飞轮中得到广泛应用^[2]。

改善无刷直流电机的转矩脉动一直是学者研究的热点,研究工作主要集中在抑制换相转矩脉动^[3-13]和消除非理想反电势(ElectroMotive Force, EMF)引起的转矩脉动^[14-15]。经典文献[3]对无刷直流电机换相转矩脉动给出了定量的分析和推导结论,但没有给出改善的方法;文献[4-7]对不同脉冲宽度调制(Pulse Width Modulation, PWM)方式对换相转矩脉动的影响进行了分析,并提出了PWM_ON_PWM调制方式来抑制换相转矩脉动,取得了较好的效果,但此种调制方式增加了位置传感器的数量,降低了系统的可靠性;文献[8]提出了一种换相期间三相配合的调制策略,达到了改善换相转矩脉动和调节换相时间的控制效果;文献[9]介绍了一种利用单个电流传感器实现无差拍电流控制的换相转矩脉

* 收稿日期:2016-01-18
基金项目:国家自然科学基金资助项目(61304036)
作者简介:冯健(1989—),男,山东德州人,博士研究生,E-mail: fengjian@nudt.edu.cn;
刘昆(通信作者),男,教授,博士,博士生导师,Email: liukun@nudt.edu.cn

动抑制策略;文献[10-12]采用 DC-DC 变换器来改变母线电压,三相桥只进行换相的控制策略,有效减小了传导区和换相区的转矩脉动,但增加了系统的复杂性,且其带宽限制了在高速场合的应用;文献[13]设计了一种简单有效的换相时间检测电路,对不同转速提出了不同的换相转矩脉动抑制策略,但没有考虑非理想反电势的影响。以上方法均只对换相转矩脉动进行抑制,没有对非理想反电势引起的力矩波动进行研究。文献[14]提出了一种新颖的非理想反电势转矩波动抑制方法,但没有对换相期间的转矩波动进行补偿,且采用的是 HPWM_LPWM 的调制方式;文献[15]以力矩电机为控制对象,采用了一种 PWM_ON_PWM 的控制方法同时对换相转矩脉动和非理想反电势转矩脉动进行改善,但该方法不适用于高速无刷直流电机。

本文针对换相期间转矩脉动提出了低速区间开通相常通,对非换相相进行 PWM 调制,高速区间对关断相进行 PWM 调制的补偿抑制策略。

1 换相力矩脉动分析

1.1 非理想反电势引起的力矩脉动

如图 1 所示,BLDCM 三相绕组为星型连接,三相电压源型逆变电路进行驱动控制。其中, U_A 、 U_B 、 U_C 分别为 A、B、C 三相绕组的端电压, i_A 、 i_B 、 i_C 分别为三相绕组电阻, R 为相电阻, L 为相电感, R_p 为能耗制动电阻, e_A 、 e_B 、 e_C 为各相反电势, U_d 为母线电压, U_{N0} 为中性点电压。

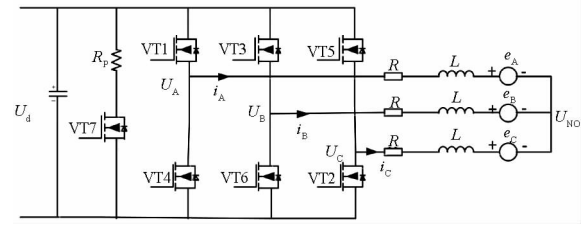


图 1 BLDCM 驱动电路拓扑框图

Fig.1 Drive circuit configuration of BLDCM

假设三相绕组分布均匀且参数一致,BLDCM 三相绕组电压平衡方程如式(1)~(3)所示。

$$U_A = Ri_A + L \frac{di_A}{dt} + e_A + U_{N0} \tag{1}$$

$$U_B = Ri_B + L \frac{di_B}{dt} + e_B + U_{N0} \tag{2}$$

$$U_C = Ri_C + L \frac{di_C}{dt} + e_C + U_{N0} \tag{3}$$

BLDCM 的电磁转矩表达式为

$$T_e = \frac{e_A i_A + e_B i_B + e_C i_C}{\omega} \tag{4}$$

式中, ω 为电机转速。对绕组为星型连接的 BLDCM 有

$$i_A + i_B + i_C = 0 \tag{5}$$

如图 2 虚线所示,理想的反电势波形为梯形波,平顶部分 120°电角度。采用两相绕组同时导通的运行方式,在反电势为平顶的部分对相应的绕组施加方波电流可以获得恒定电磁转矩。

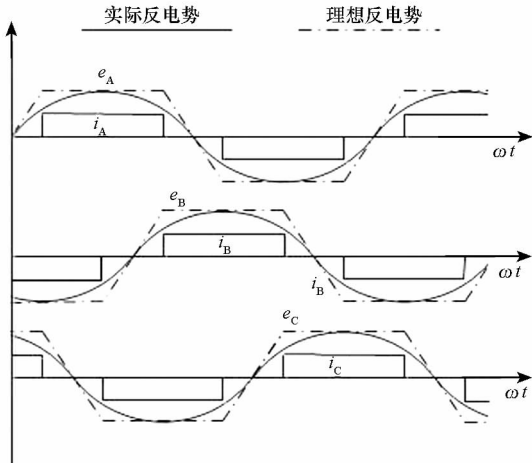


图 2 BLDCM 反电势和绕组电流时序图

Fig.2 Back EMF and phase current sequence in time of BLDCM

但高速无刷直流电机本体由于设计和制造方面的原因,很难做到反电势为平顶宽度 120°电角度的梯形波,实际上绕组反电势更接近正弦波,如图 2 实线所示。按照传统控制方法,对绕组施加恒定电流值,由式(6)知,会引起电机输出电磁力矩脉动。

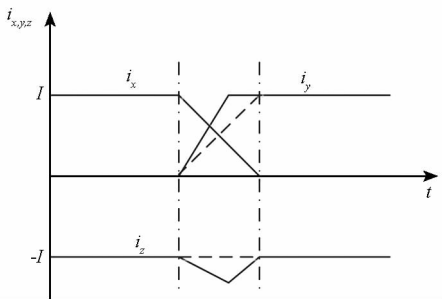
$$T_e = \frac{2EI}{\omega} = k_T I \tag{6}$$

式中, E 和 I 分别为绕组反电势和电流幅值, k_T 为力矩系数。

1.2 换相转矩脉动

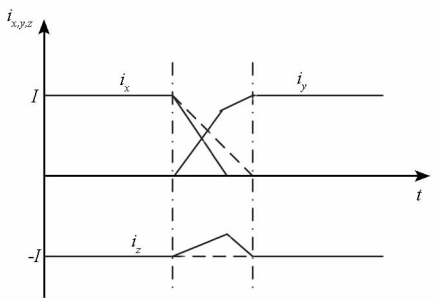
BLDCM 采用三相六状态控制,换相过程中,由于绕组电感的存在,绕组电流存在上升和下降时间,且变化率不一致。理论上,电机低速运行($U_d > 4E$)时,导通相电流变化率大于关断相,如图 3(a)所示,换相过程中电磁转矩增大;电机高速运行($U_d < 4E$)时,导通相电流变化率小于关断相^[3],如图 3(b)所示,换相过程中电磁转矩减小。导通相和关断相电流变化不一致导致非换相相电流波动,进而引起电机电磁转矩出现增大或减小,甚至幅值高达 50%^[3]。解决换相力矩波动的基本思想是匹配合换相过程中导通相和关断相的电流变

化率,使非换相相电流保持恒定。由于增大变化较慢相的电流变化率需要提高换相过程中的母线电压,大大增加电路的复杂性,故本文针对不同转速区的各相电流变化特点,采取减小电流变化较快相的电流变化率的补偿控制策略来抑制换相转矩脉动。



(a) 低速区间换相电流

(a) Commutation currents waveform in low-speed range



(b) 高速区间换相电流

(b) Commutation currents waveform in high-speed range

图3 换相过程电流变化

Fig. 3 Waveforms of commutation currents

2 转矩脉动抑制方法

2.1 传导区转矩脉动抑制

非换相区转矩脉动主要由非理想反电势引起,本文采用 HPWM_LON 的调制方式。以 A、C 相导通为例,电流从 A 相流入, C 相流出,则 MOSFET VT1 进行 PWM 调制, MOSFET VT2 常通。由基尔霍夫定律有

$$\begin{cases} SU_d = Ri_A + L \frac{di_A}{dt} + e_A + U_{NO} \\ 0 = Ri_C + L \frac{di_C}{dt} + e_C + U_{NO} \end{cases} \quad (7)$$

中性点对地电压为

$$U_{NO} = \frac{SU_d - (e_A + e_C)}{2} \quad (8)$$

开关管 VT1 开通时 $S = 1$, 关断时 $S = 0$ 。在一个 PWM 周期 T_s 中,在 DT_s 区间内 S 保持为 1, 在 $(1-D)T_s$ 区间内 S 保持为 0 (D 为占空化)。由状态空间平均法有

$$U_{NO} = \frac{DU_d - (e_A + e_C)}{2} \quad (9)$$

由于采样频率 (20 kHz) 较高,在每一个采样周期内各相反电势取为一恒定值。采样周期远小于 BLDCM 电磁时间常数,故忽略绕组电阻的影响。在一个 PWM 周期内, A、C 两相的电流变化率为

$$\left[\frac{di_A}{dt} \right] = \frac{DU_d + e_C - e_A}{2L} \quad (10)$$

$$\left[\frac{di_C}{dt} \right] = \frac{e_A - e_C - DU_d}{2L} \quad (11)$$

当前采样周期内 A 相绕组电流为 $i_A(k)$, 则下一采样周期 A 相绕组电流为

$$i_A(k+1) = \frac{D(k)U_d(k) + e_C(k) - e_A(k)}{2L} T_s + i_A(k) \quad (12)$$

令 $i_A(k+1) = I_{ref}$, 得下个采样周期内 PWM 占空比为

$$D(k) = \frac{2L[I_{ref} - i_A(k)]}{T_s U_d(k)} + \frac{e_A(k) - e_C(k)}{U_d(k)} \quad (13)$$

2.2 低速换相区

BLDCM 每隔 60° 电角度进行一次换相,以获取最大电磁力矩。由于绕组电感的存在,换相期间三相绕组均有电流通过。根据设计的低速区间的换相转矩波动补偿策略,以 A + C - 过渡到 B + C - 为例。换相过程中,关断 VT1,保持 VT3 导通,对 VT2 进行 PWM 调制。由基尔霍夫电压定律有

$$\begin{cases} 0 = Ri_A + L \frac{di_A}{dt} + e_A + U_{NO} \\ U_d = Ri_B + L \frac{di_B}{dt} + e_B + U_{NO} \\ (1-S)U_d = Ri_C + L \frac{di_C}{dt} + e_C + U_{NO} \end{cases} \quad (14)$$

此时中性点电压为

$$U_{NO} = \frac{(2-S)U_d - (e_A + e_B + e_C)}{3} \quad (15)$$

开关管 VT2 开通时 $S = 1$, 关断时 $S = 0$ 。在一个 PWM 周期 T_s 中,在 $D_L T_s$ 区间内 S 保持为 1, 在 $(1-D_L)T_s$ 区间内保持为 0 (D_L 表示低速区间的占空比)。由状态空间平均法有

$$U_{NO} = \frac{(2-D_L)U_d - (e_A + e_B + e_C)}{3} \quad (16)$$

换相期间,在一个 PWM 周期内,关断相 A 和开通相 B 的电流变化率分别为

$$\left[\frac{di_A}{dt} \right] = \frac{(D_L - 2)U_d + e_B + e_C - 2e_A - 3I_A R}{3L} \quad (17)$$

$$\overline{\left[\frac{di_B}{dt}\right]} = \frac{(D_L + 1)U_d + e_A + e_C - 2e_B - 3\overline{i_B}R}{3L} \quad (18)$$

要保持非换相 C 相电流不变,令 A 相电流变化率和 B 相电流变化率绝对值相等,即

$$-\overline{\left[\frac{di_A}{dt}\right]} = \overline{\left[\frac{di_B}{dt}\right]} \quad (19)$$

将式(17)和式(18)代入式(19)得低速区间内,换相期间每个 PWM 周期内的占空比为

$$D_L = \frac{U_d(k) + e_A(k) + e_B(k) - 2e_C(k) + 3IR}{2U_d(k)} \quad (20)$$

2.3 高速换相区

电机运行在高速区间,开通相的电流变化率较慢,对关断相进行 PWM 调制,以控制关断相的电流变化率,使两者相等。换相仍以 A + C - 过渡到 B + C - 为例,对 VT1 进行 PWM 调制,保持 VT2、VT3 导通。同样,由基尔霍夫电压定律有

$$\begin{cases} SU_d = Ri_A + L \frac{di_A}{dt} + e_A + U_{NO} \\ U_d = Ri_B + L \frac{di_B}{dt} + e_B + U_{NO} \\ 0 = Ri_C + L \frac{di_C}{dt} + e_C + U_{NO} \end{cases} \quad (21)$$

中性点电压为

$$U_{NO} = \frac{(1+S)U_d - (e_A + e_B + e_C)}{3} \quad (22)$$

$$U_{NO} = \frac{(1+D_H)U_d - (e_A + e_B + e_C)}{3} \quad (23)$$

其中, D_H 为高速换向期间的占空比。在一个 PWM 周期内,关断相 A 和导通相 B 的平均电流变化率分别为

$$\overline{\left[\frac{di_A}{dt}\right]} = \frac{e_B + e_C - 2e_A + (2D_H - 1)U_d - 3\overline{i_A}R}{3L} \quad (24)$$

$$\overline{\left[\frac{di_B}{dt}\right]} = \frac{e_A + e_C - 2e_B + (2 - D_H)U_d - 3\overline{i_B}R}{3L} \quad (25)$$

要保持非换相 C 相电流不变,令 A 相电流变化率和 B 相电流变化率绝对值相等,即

$$-\overline{\left[\frac{di_A}{dt}\right]} = \overline{\left[\frac{di_B}{dt}\right]} \quad (26)$$

同样,将式(24)和式(25)代入式(26)得高速区间内,换相期间每个 PWM 周期内的占空比为

$$D_H = \frac{e_A(k) + e_B(k) - 2e_C(k) - U_d(k) + 3IR}{U_d(k)} \quad (27)$$

3 换相时间计算

要实现换相转矩波动的精确补偿,需确定换相控制过程持续时间。假设完成换相过程需要 n 个 PWM 周期 T_s ,则换相时间为 nT_s 。换相开始时刻绕组工作电流 I 和换相时间满足关系式

$$nT_s \cdot \overline{\left[\frac{di}{dt}\right]} = I \quad (28)$$

由于换相时间较短,计算换相时间时,认为绕组反电势保持为换相开始时刻的反电势值。此时有 $e_A = e_B = -e_C = E$ 和 $D_s U_d = 2(IR + E)$, D_s 为换向开始时刻的占空比值,则低速和高速区间,每个采样周期内 PWM 占空比可分别进一步简化为

$$D_L = \frac{1}{2} + \frac{2E + 3IR}{2U_d} = \frac{1}{2} + D_s - \frac{IR}{2U_d} \quad (29)$$

$$D_H = \frac{4E - U_d + 3IR}{U_d} = 2D_s - 1 - \frac{IR}{U_d} \quad (30)$$

将上两式分别代入式(24)和式(25),得低速区间和中高速区间的 n 值分别为

$$n_L = \frac{2IL}{U_d T_s} \quad (31)$$

$$n_H = \frac{2IL}{[2U_d(1 - D_s) + IR]T_s} \quad (32)$$

4 反电势估算

电机的反电势与转速成正比^[4],是转速和转子角位置的函数。因此通过离线检测不同转速的反电势数据,构造反电势波形函数,能够获取准确的反电势波形。六状态导通模式下,在每个导通区的相反电势波形可通过式(33)和式(34)来确定。式中, $\omega(k)$ 和 $\varphi(k)$ 分别为当前采样周期的转速值和转子位置电角度值, $f[\varphi(k)]$ 为离线检测得到的反电势波形函数。各相导通区反电势波形函数如表 1 所示,将其存入查找表中。

$$e_x(k) = \omega(k)f[\varphi(k)] \quad (33)$$

$$e_y(k) = \omega(k)f\left[\frac{\pi}{3} - \varphi(k)\right] \quad (34)$$

以换相开始时刻为参考点,即 $\omega t = 0$,转子位置电角度值 $\varphi(k)$ 可由式(35)计算得到,其中 $0 \leq \varphi(k) \leq \pi/3$ 。

$$\varphi(k) = \varphi(k-1) + \omega(k)T_s, k \geq 2 \quad (35)$$

控制器检测到换相标志后,将 k 值归 1,令 $\varphi(1) = \omega(1)T_r$,其中 $\omega(1)$ 为当前采样周期的转速值, T_r 为被换相时刻分割的当前采样周期残余值。

分别在转速为 250 r/min, 500 r/min, 750 r/min, 1000 r/min, 1250 r/min 和 1500 r/min

表 1 导通区各相反电势函数

Tab.1 Phase back EMF shape function in conduction region

霍尔状态	$e_A(k)$	$e_B(k)$	$e_C(k)$
100	$e_x(k)$	$-e_y(k)$	—
110	$e_y(k)$	—	$-e_x(k)$
010	—	$e_x(k)$	$-e_y(k)$
011	$-e_x(k)$	$e_y(k)$	—
001	$-e_y(k)$	—	$e_x(k)$
100	—	$-e_x(k)$	$e_y(k)$

时,对表 1 描述的样机反电势进行离线检测,检测数据如图 4、图 5 所示。以线反电势过零点为基准,图 4 给出了每个电周期内电角度 $60^{\circ}, 70^{\circ}, 80^{\circ}, 90^{\circ}$ 位置处的转速 - 反电势检测曲线,各处反电势均具有较好的线性度,验证了之前的假设。图 5(a) 为 1500 r/min 转速下线反电势的离线检测数据,细线为直接测量数据,粗线为测量数据经过最小二乘滤波得到的曲线。从图中曲线可以直观地看出,反电势波形较理想的平顶梯形波相差较大。图 5(b) 中粗线为转速 2000 r/min 时,利用本文方法得到的估算反电势值,细线为直接测量数据。由图表明,估算得到的反电势数据能十分准确地反应实际反电势值,证明了本文反电势估算方法的有效性和准确性。

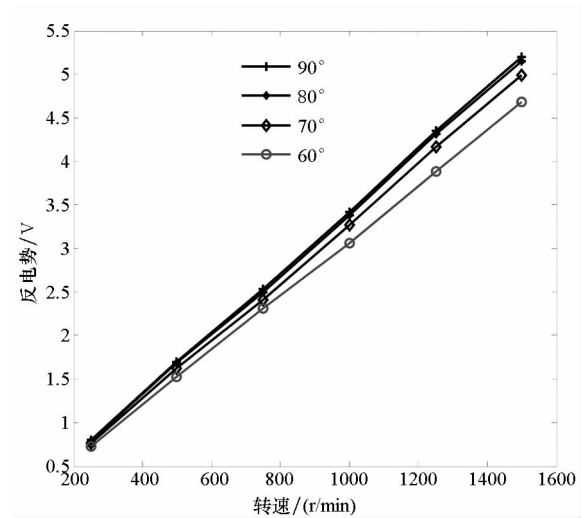
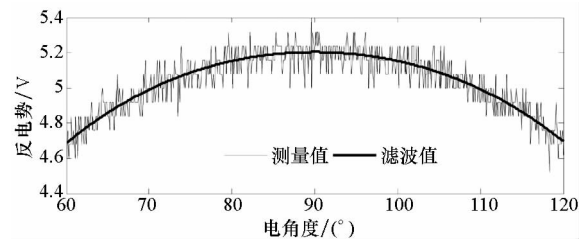


图 4 不同转速线反电势曲线图

Fig.4 Back line-EMF with different motor speed

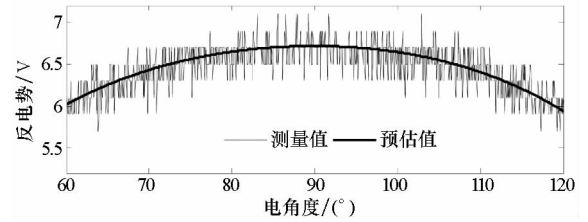
5 实验验证及分析

为验证所提控制方法,在磁悬浮飞轮原理样机上进行了实验。磁悬浮原理样机参数如表 2 所示,样机以 TMS320C6731 DSP 和 Spartan



(a) 1500 r/min 线反电势检测值与滤波值

(a) Measured values and filtered values of back line-EMF at 1500 r/min



(b) 2000 r/min 线反电势测量值与估计值

(b) Measured values and estimated values of back line-EMF at 2000 r/min

图 5 线反电势估算曲线

Fig.5 Estimated back line-EMF

XC3S1000 FPGA 为控制器。采样电流值和位置信号经现场可编程门阵列 (Field-Programmable Gate Array, FPGA) 处理后,数字信号处理器 (Digital Signal Processor, DSP) 读取数据进行运算,FPGA 根据数据处理结果生成换相逻辑和 PWM 调制信号。

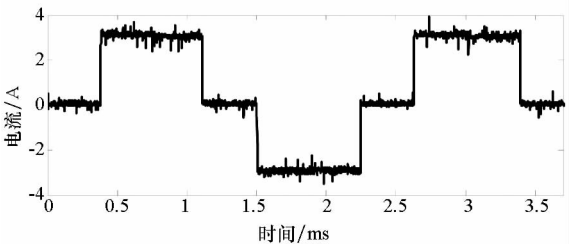
表 2 样机参数

Tab.2 BLDCM ratings

参数	数值
额定电压/V	28
最大输出力矩/(N·m)	>0.1
极对数	8
力矩系数/(N·m/A)	0.017
相电阻/ Ω	0.47
相电感/mH	0.18

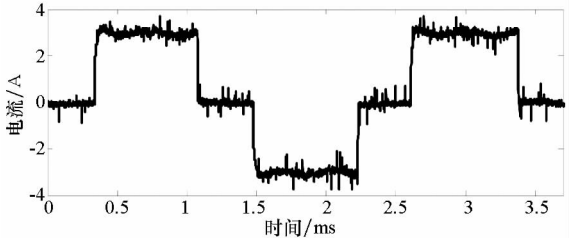
根据本文设计的补偿控制策略,分别对非理想反电势、换相力矩脉动抑制进行了实验。图 6(a) 为未对非理想反电势进行补偿、传统控制方式下的绕组电流波形,电流为一恒定值。图 6(b) 为补偿后的电流波形,电流幅值随反电势幅值改变实时变化。

图 7 为对非理想反电势进行补偿前后 BLDCM 输出电磁力矩曲线。补偿前,如图 7(a) 所示,输出



(a) 补偿前电流波形

(a) Current waveform without compensation



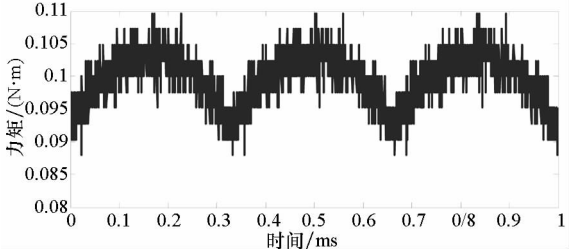
(b) 补偿后电流波形

(b) Current waveform with compensation

图 6 反电势补偿电流波形

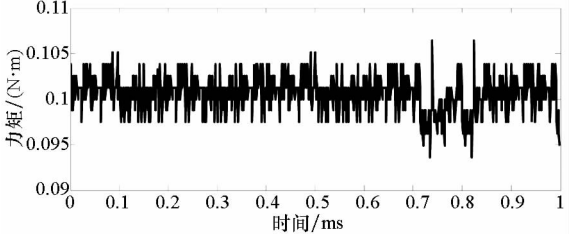
Fig. 6 Current waveform with back EMF compensation

力矩输出值略小于指令电流 $0.1\text{ N}\cdot\text{m}$, 且力矩脉动达 17%; 图 7(b) 为补偿后力矩波形, 能够准确复现力矩指令, 力矩波动减小为 5%。



(a) 补偿前输出力矩波形

(a) Torque waveform without compensation



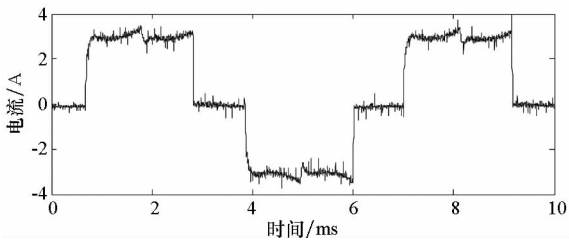
(b) 补偿后输出力矩波形

(b) Torque waveform with compensation

图 7 反电势补偿前后输出力矩波形

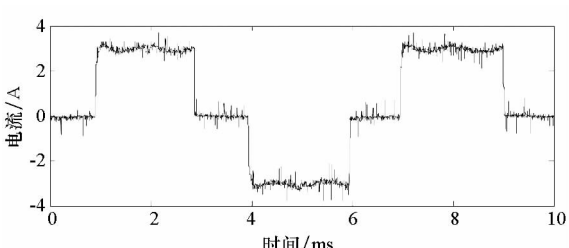
Fig. 7 Torque waveform with back EMF compensation

图 8、图 9 分别为低速区间换相力矩补偿前后电流和力矩波形。图 8(b) 与图 8(a) 相比, 消除了非换相相电流在低速区间因换相引起的增大, 从而使换相力矩波动从 19% 减小为 5% (如图 9 所示)。



(a) 换向补偿前电流波形

(a) Current waveform without commutation compensation

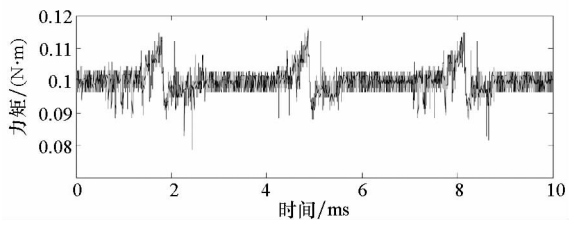


(b) 换向补偿后电流波形

(b) Current waveform with commutation compensation

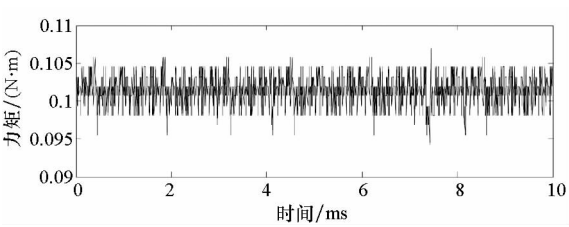
图 8 低速区间换相补偿电流波形

Fig. 8 Current waveform of commutation compensation in low-speed range



(a) 换向补偿前输出力矩波形

(a) Torque waveform without commutation compensation



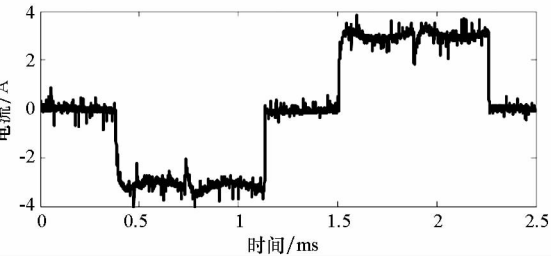
(b) 换向补偿后输出力矩波形

(b) Torque waveform with commutation compensation

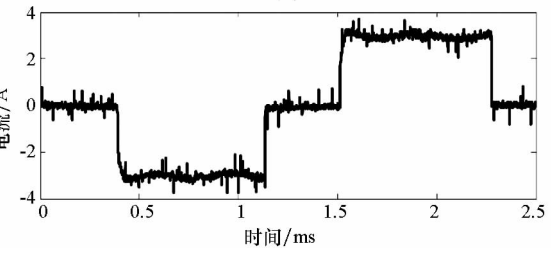
图 9 低速区间换相补偿力矩波形

Fig. 9 Torque waveform of commutation compensation in low-speed range

图 10、图 11 分别为高速区间换相力矩补偿前后电流和力矩波形。图 10(b) 与图 10(a) 相比, 消除了非换相相电流在高速区间因换相引起的减小, 从而使换相力矩波动从 31% 减小为 11% (如图 11 所示)。

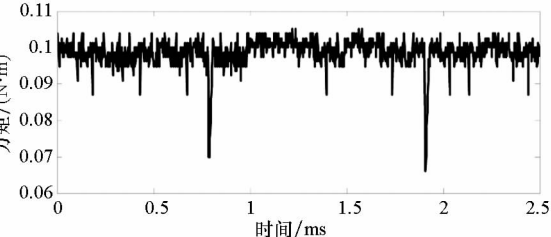


(a) 换向补偿前电流波形
(a) Current waveform without commutation compensation

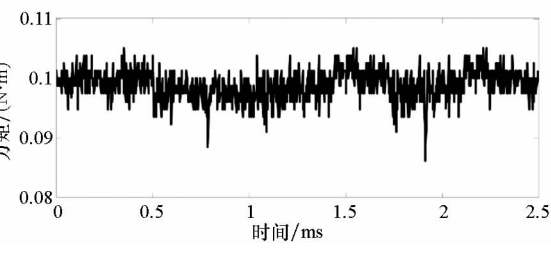


(b) 换向补偿后电流波形
(b) Current waveform with commutation compensation

图 10 高速区间换相补偿电流波形
Fig. 10 Current waveform of commutation compensation during high-speed range



(a) 换向补偿前输出力矩波形
(a) Torque waveform without commutation compensation



(b) 换向补偿后输出力矩波形
(b) Torque waveform with commutation compensation

图 11 高速区间换相补偿力矩波形
Fig. 11 Torque waveform of commutation compensation during high-speed range

6 结论

对磁悬浮飞轮用永磁无刷直流电机传统控制方式中因非理想反电势和换相存在的转矩脉动现象进行了分析,针对换相转矩脉动分别提出了低速区间开通相常通,对非换相相进行

PWM 调制,高速区间对关断相进行 PWM 调制的补偿策略。对比实验表明该策略取得了很好的转矩脉动抑制效果,从而验证了该补偿控制策略的有效性,为磁悬浮飞轮高精度力矩控制提供了新的方法。

参考文献 (References)

[1] Chou M C, Liaw C M, Chien S B, et al. Robust current and torque controls for PMSM driven satellite reaction wheel[J]. IEEE Transactions on Aerospace Electronic System, 2011, 47(1): 58 – 74.

[2] Varatharajoo R, Fasoulas S. The combined energy and attitude control system for small satellites-Earth observation missions [J]. Acta Astronautica, 2005, 56 (1/2): 251 – 259.

[3] Carlson R, Lajoie-Mazenc M, Fagundes J C. Analysis of torque ripple due to phase commutation in brushless DC machines[J]. IEEE Transactions on Industry Applications, 1992, 28(3): 632 – 638.

[4] 周美兰, 高肇明, 吴晓刚, 等. 五种 PWM 方式对直流无刷电机系统换相转矩脉动的影响[J]. 电机与控制学报, 2013, 17(7): 15 – 21.

ZHOU Meilan, GAO Zhaoming, WU Xiaogang, et al. Influence of five kinds of PWM on commutation torque ripples in BLDCM control system [J]. Electric Machines and Control, 2013, 17(7): 15 – 21. (in Chinese)

[5] 韦鲲, 胡长生, 张仲超. 一种新的消除无刷直流电机非导通续流的 PWM 调制方式[J]. 中国电机工程学报, 2005, 25(7): 104 – 108.

WEI Kun, HU Changsheng, ZHANG Zhongchao. A novel PWM scheme to eliminate the diode freewheeling of the inactive phase in BLDC motor[J]. Proceedings of the CSEE, 2005, 25(7): 104 – 108. (in Chinese)

[6] Wei K, Hu C S, Zhang Z C. A novel commutation torque ripple suppression scheme in BLDCM by sensing the DC current [C]//Proceedings of 36th IEEE Power Electronics Conference, 2005: 1259 – 1263.

[7] Meng G W, Hao X, Li H S. Commutation torque ripple reduction in BLDC motor using PWM_ON_PWM mode[C]// Proceedings of International Conference on Electrical Machines and Systems, 2009: 1 – 6.

[8] 石坚, 李铁才. 一种消除无刷直流电动机换相转矩脉动的 PWM 调制策略[J]. 中国电机工程学报, 2012, 32(24): 110 – 116.

SHI Jian, LI Tiecai. A PWM strategy eliminate commutation torque ripple of brushless DC motor[J]. Proceedings of the CSEE, 2012, 32(24): 110 – 116. (in Chinese)

[9] Song J H, Choy I. Commutation torque ripple reduction in brushless DC motor drives using a single DC current sensor[J]. IEEE Transactions on Power Electronics, 2004, 19(2): 312 – 319.

[10] Shi T N, Guo T T, Song P, et al. A new approach of minimizing commutation torque ripple for brushless DC motor based on DC-DC converter [J]. IEEE Transactions on Industrial Electronics, 2010, 57(10): 3483 – 3490.

相控阵天线阵面两级备件优化配置模型*

王永攀^{1,2}, 杨江平¹, 张宇³, 侯晓东¹
(1. 空军预警学院 陆基预警装备系, 湖北 武汉 430019;
2. 中国人民解放军 93502 部队, 内蒙古 呼和浩特 010051;
3. 湖北工业大学 电气与电子工程学院, 湖北 武汉 430068)

摘要:针对相控阵天线阵面备件配置存在的冗余性强、批量送修、多级维修等现实问题,综合考虑备件费用、维修能力以及库存策略之间的关系,建立了基于定期补给的两级备件优化配置模型。给出了系统的故障件维修周转过程和维修备件的定期补给过程,在分析备件、库存、维修能力之间关系的基础上,结合成批到达的排队理论,建立了系统的供应可用度模型。以备件配置费用最小为目标、以系统供应可用度为约束条件,建立了系统的备件优化配置模型,并通过边际效益分析法对模型进行了求解。通过算例仿真与分析对模型进行了验证。结果表明:构建的备件配置能够较好地解决相控阵天线阵面的备件配置问题,具有一定的优越性。

关键词:相控阵天线;两级维修;备件; K/N 系统;批量送修

中图分类号:TN95;N94 **文献标志码:**A **文章编号:**1001-2486(2017)03-172-07

Optimal configuration model of spare parts for phased array antenna under two-echelon maintenance supply

WANG Yongpan^{1,2}, YANG Jiangping¹, ZHANG Yu³, HOU Xiaodong¹
(1. Land-based Early Warning Equipment Department, Air Force Early Warning Academy, Wuhan 430019, China;
2. The PLA Unit 93502, Hohhot 010051, China;
3. College of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, China)

Abstract: Three problems are normally found in the spare parts configuration of phased array antenna, namely, the strong redundancy, the batch delivery maintenance and the multi-echelon maintenance. Aiming at these problems, through analyzing the relations among spare parts cost, repair capacity and inventory strategies, an optimal configuration model was established based on the periodic review strategies. Firstly, the repair circulation process of fault component and the periodic supply process of maintenance spare parts were given, and then the system supply availability model was built by analyzing the relations among spare parts, inventories and repair capacity, and the batch arrival queuing theory was also used. Secondly, an optimal configuration model of spare parts was built, which takes minimum spare parts costs as the object and the system availability as the subject. Next, solution algorithm based on the margin analysis theory to the model was also given. Finally, simulations and analysis of an instance were conducted to verify the proposed model, and results show that the model can solve the spare parts allocation problems of phased array antenna well, and has a high superiority.

Key words: phased array antenna; two-echelon maintenance; spare parts; K/N systems; batch delivery maintenance

大型相控阵雷达 (Large-scale Phased Array Radar, LPAR) 在国家战略预警尤其是反导预警作战中作用重大,地位特殊。作为相控阵雷达的主要组成部分,天线阵面分系统的维修备件配置问题已成为部队和科研院所研究和关注的重点。当前,部队的普遍做法是采用国军标 GJB 4355《备件供应规划要求》中的单项备件配置方法^[1]。然而,从当前部队的实际状况来看,该方法暴露出以下问题:①不能兼顾备件配置费用、维修能力等限制条件与系统可用度的关系;②不能兼顾系统冗余设计对系统备件配置水平的影响;③不能较好地分配基层级和基地级备件的库存量。上述三个典型的问题已成为影响相控阵雷达军事和经济效益的重要因素,急需探求合理的解决方法。

由于天线阵面组件众多且采用冗余设计,因此,许多研究人员将天线阵面看作一个 K/N 系统

* 收稿日期:2016-01-11
基金项目:军队科研资助项目(KJ2014023200B11145);博士研究生专项资助项目(2014JY546)
作者简介:王永攀(1987—),男,河北保定人,博士研究生,Email:wypaning@163.com;
杨江平(通信作者),男,教授,博士,博士生导师,Email:yjp_wh@163.com

来进行研究,并取得了一定的研究成果。如:文献[2]研究了多级冗余的 $K/N(G)$ 系统(即系统是 K/N 冗余系统,子系统是 $1/m(F)$ 冗余系统)的备件配置问题;文献[3]针对两级(系统级、部件级)均为 $K/N(G)$ 结构的冗余系统备件配置问题进行了研究;文献[4]研究了考虑报废的系统级 K/N 冷备份冗余系统的备件配置问题。上述文献在一定程度上解决了当前存在的问题,但是仍存在一些局限性。如这些研究均认为系统存在单个部件故障后立即进行送修,即单件送修的情况;而在实际应用中,对于采用冗余设计的相控阵雷达天线阵面而言,批量送修的问题客观存在,需进一步开展研究。为此,部分学者开始研究批量送修的情况。所谓批量送修是指对系统进行维修时,系统中的故障件已达到一定数量,需成批地将故障件送至维修点进行维修。已有研究表明:对于批量送修的 K/N 系统而言,备件数目与系统的维修策略以及维修能力之间相互作用、相互影响^[5]。为此,部分学者综合考虑三者之间的关系,从故障件维修的角度出发,通过求取系统的使用可用度,综合考虑系统备件配置费用的影响,构建了考虑批量送修的 K/N 系统备件配置模型^[6-9]。这些研究虽然解决了批量送修的问题,但是带来了新的问题,主要表现在两方面:①由于批量送修条件下经典的 $(S-1, S)$ 库存策略不再适用,因此,这些研究仅仅从维修的角度来研究 K/N 系统的备件配置问题;而从库存的角度出发来研究考虑批量送修的 K/N 系统的备件配置问题,更符合用户的客观需求。②仅仅考虑了单级维修条件下 K/N 系统的备件配置问题,而在实际应用中,多级维修条件下的备件问题更贴近实际。为此,仍需要进一步开展相应的研究。本文研究了基于定期补给库存策略的两级维修条件下天线阵面的备件优化配置问题。重点研究了车间可更换单元(Shop Replacement Unit, SRU)层级的备件配置问题。构建了系统的供应可用度模型,并以备件费用最小为优化目标、以供应可用度为约束条件,构建了天线阵面的备件优化配置模型,通过边际效益分析方法对模型进行了求解。

1 问题描述与假设

为了便于研究,将天线阵面看作一个由 N 个现场可更换单元(Line Replacement Unit, LRU)组成的 K/N 系统,以下 K/N 系统均指天线阵面。设 K/N 系统中每个 LRU 由 M 个 SRU_i 串联组成,当

任意一个 SRU_i 故障时,对应的 LRU 故障;系统正常工作时,要求系统中至少有 K 个 LRU 正常工作, K/N 系统的具体结构如图 1 所示。

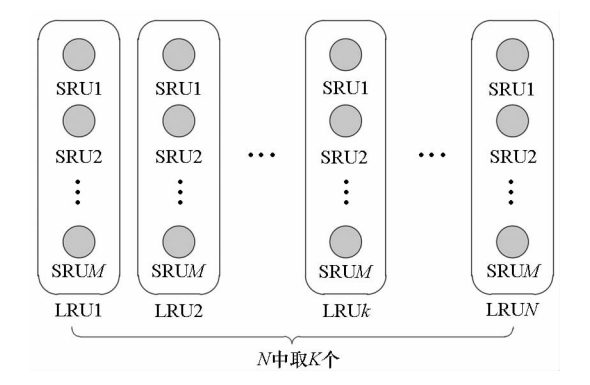


图 1 K/N 系统结构图
Fig.1 Structure of K/N system

考虑到 K/N 系统的冗余性,在工程应用中一般对系统进行预防性维修。当系统运行到预防性维修阈值时,认为系统故障,需对系统进行停机维修。维修时,基层级负责更换系统中故障的 LRU,其中,LRU 备件通常由基层级维修点修复得到。由于 LRU 的故障是由 SRU_i 引起的,在定位故障后,通常进行 SRU_i 换件维修。如果有对应的 SRU_i 备件,则通过更换 LRU 中故障的 SRU_i 修复 LRU;如果没有备件,则发生一次备件短缺,需等待一定时间才能补充。在基层级更换下来的 SRU_i 故障件被成批地送到基地级进行维修,在进行定期补给时,如果基地级有 SRU_i 备件,则直接补给基层级库存,如果基地级没有 SRU_i 备件,则发生一次 SRU_i 备件短缺,需等待基地级维修点修复 SRU_i 备件。在基地级库存发生额外的 SRU_i 备件需求时,通过采购的方式进行补充。图 2 给出了故障 SRU_i 的维修周转过程。

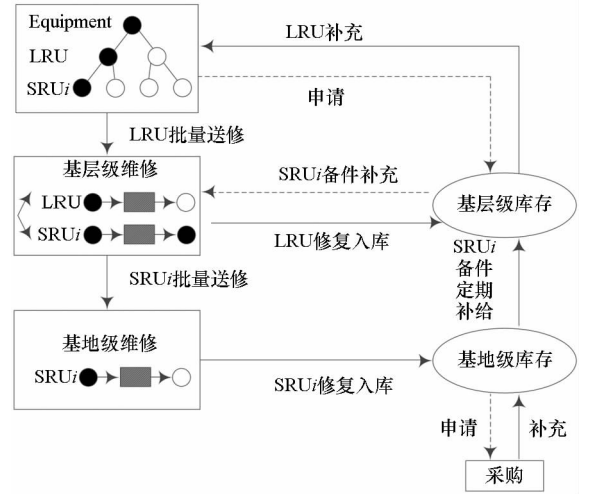


图 2 故障件维修周转过程
Fig.2 Repair circulation process of fault component

图 3 给出了系统维修备件定期补给过程。

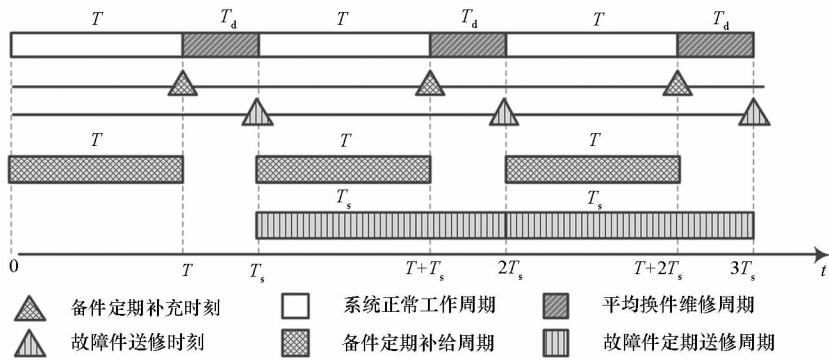


图 3 维修备件定期补给过程

Fig. 3 Periodic supply process of maintenance spare parts

图 3 中： T 表示系统正常工作的周期， T_d 为系统的平均换件维修周期， T_s 为系统中故障件的送修周期， $T_s = T + T_d$ ，其中，系统的备件定期补给周期与系统的正常工作周期保持一致。设从 0 时刻开始，经过时间 T ，系统需停机进行维修，此时，备件已从基地级送至基层级进行补给；经过 T_d 时间，即在 T_s 时刻，系统完成修复，此时，将未修复的故障件送至基地级进行维修；再过时间 T ，即在 $T + T_s$ 时刻，将备件从基地级送至基层级进行补充；如此，反复执行，此过程即为备件的定期补给过程。

鉴于 K/N 系统的备件配置结果与系统维修策略及库存策略有很大的关系，为便于开展研究，需做出如下假设。

- 1) 系统中所有 LRU 服从参数为 λ_{LRU} 的指数分布，所有 SRU i 服从参数为 λ_{SRU_i} 的指数分布；
- 2) 所有故障件均能得到修复，维修过程中不存在报废且修复如新，可作为备件使用；
- 3) 故障件采取两级维修体制，即基层级维修和基地级维修，基层级只负责对故障 LRU 进行 SRU i 换件维修，基地级负责故障 SRU i 的维修且维修能力有限；
- 4) SRU i 维修渠道不同，维修时均遵循排队原则，维修工作相互独立且同时进行，一个维修渠道只能同时修理一个故障件，SRU i 的维修渠道总数小于 SRU i 故障件的数目；
- 5) 基层级采取 (T, S) 定期补给库存策略，即以固定周期 T 对基层级库存进行定期补给，将基层级库存补充到 S ；
- 6) 不考虑 LRU 中除 SRU i 之外的组成部分对 LRU 故障造成的影响。

2 系统备件优化配置模型

令 $i > 0$ 代表 SRU；令 $j = 0$ 代表基地级， $j = 1$

代表基层级。如 $S_{ji} = S_{03}$ 代表基地级 SRU3 备件的库存量， $S_{ij} = S_{11}$ 代表基层级 SRU1 备件的库存量。结合故障件的维修周转过程可知，SRU i 的备件库存由基层级库存 S_{1i} 和基地级库存 S_{0i} 两部分组成，为保证系统的正常运行，基地级必须有充足的备件来满足基层级的库存。设备件 SRU i 的需求率为 m_i ，则 m_i 的计算公式^[10]为：

$$m_i = \lambda_{SRU_i} \cdot N \cdot U \cdot R \tag{1}$$

式中： N 为 SRU i 总数； U 为产品利用率因子； R 为拆卸率因子，取 $U = 1, R = 1$ 。

系统的工作周期可分为两个阶段，包括系统正常运行时间 $T = ET$ 和 LRU 换件维修时间 $T_d = ED$ 。根据 K/N 系统定义可知，当系统中故障件数目达到 $m_{\max} = N - K + 1$ 个时，整个系统会停止工作。则系统的平均工作时间可表示为：

$$ET = \sum_{m=0}^{N-K+1} \frac{1}{(N-m)\lambda_{LRU}} \tag{2}$$

设有 c_r 个 LRU 换件维修人员，换件维修率为 μ_r ，则可得到 LRU 换件维修的时间为：

$$ED = \frac{m_{\max}}{c_r \mu_r} \tag{3}$$

在 T 时间段内，系统正常运行，会出现故障的 LRU，因此，也会出现 SRU i 故障。由于 SRU i 的故障是随机发生的，因此，SRU i 备件的需求服从泊松过程。可求得系统对 SRU i 备件的平均需求为 $m_i T$ 时， T 时间内发生 x_i 次需求的概率为：

$$p(x_i | m_i T) = (m_i T)^{x_i} \cdot \frac{e^{-m_i T}}{x_i!} \tag{4}$$

SRU i 故障后，被成批地送至基地级维修点进行维修。设 m_{mi} 为系统中有 m_{\max} 个 LRU 故障时对应的 SRU i 的数目，为了保障备件的充足，其值应满足：

$$m_{mi} = m_{\max} \frac{\lambda_{SRU_i}}{\lambda_{LRU}} \tag{5}$$

设SRU*i*的维修渠道为*c_i*个,则故障件的维修过程可以看作一个到达率为Λ、维修率为μ_{*i*}、具有*m_{mi}*个顾客和*c_i*个服务台的*M^{mmi}/M/c_i*排队系统。根据维修备件的定期补给过程可知,故障件到达基地级的周期*T_s* = *T* + *T_d*。其中,在*T_d*时间段内,系统处于停机维修状态,在此期间没有新的

故障LRU产生。因此,可以得到故障件的到达率为:

$$\Lambda = \frac{1}{T_s} = \frac{1}{T + T_d} \tag{6}$$

进一步可画出*M^{mmi}/M/c_i*排队系统的状态流程图如图4所示。

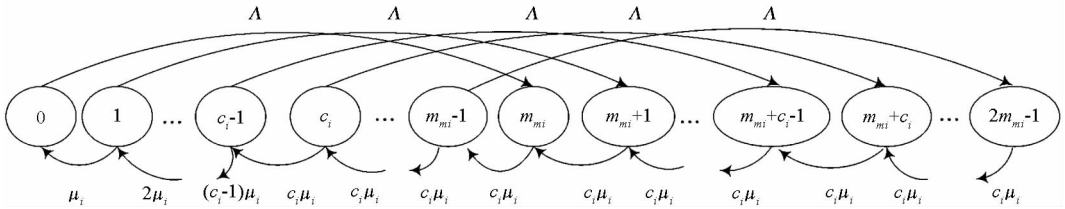


图4 状态流程图

Fig.4 State flow chart

根据状态流图,可列出平衡条件下的K氏代数方程。

$$\begin{cases} p_{y_i}(\rho_i + y_i) = p_{y_i+1}(y_i + 1) & 0 < y_i \leq c \\ p_{y_i}(\rho_i + c_i) = p_{y_i+1}c_i & c_i < y_i \leq m_{mi} \\ p_{y_i}c_i = p_{y_i-m_{mi}}\rho_i + p_{y_i+1}c & m_i < y_i \leq 2m_{mi} - 1 \end{cases} \tag{7}$$

设存在比例系数ρ_{*i*} = Λ/μ_{*i*},则系统中恰有*y_i*个SRU*i*的稳态概率,其可表示为:

$$p_i(y_i) = \begin{cases} p_{i0} & y_i = 0 \\ p_{i0}\alpha_i(y_i) & 0 < y_i \leq c_i \\ p_{i0}\alpha_i(c_i)\beta_i(y_i) & c_i < y_i \leq m_i \\ p_{i0}\alpha_i(c_i)\beta_i(m_i) - \gamma_i(y_i) & m_i < y_i \leq 2m_i - 1 \end{cases} \tag{8}$$

式中:α_{*i*}(*y_i*),β_{*i*}(*y_i*),γ_{*i*}(*y_i*)分别为在修或待修部件数*y_i*的函数。其表达式为:

$$\begin{cases} \alpha_i(y_i) = \frac{\prod_{j=0}^{y_i-1} (\rho_i + j)}{y_i!} \\ \beta_i(y_i) = \left(\frac{\rho_i + c_i}{c_i} \right)^{y_i - c_i} \\ \gamma_i(y_i) = \sum_{j=0}^{y_i - m_{mi} - 1} \frac{p(j)\rho_i}{c_i} \end{cases} \tag{9}$$

由于系统中所有稳态概率之和为1,令*p(j)*表示提取系数*p_{i0}*后*p_i(*y_i*)的值,则可以得到系统的稳态概率为:*

$$p_{i0} = \frac{1}{\xi} \tag{10}$$

$$\text{令 } \xi = 1 + \sum_{y_i=1}^{c_i} \alpha_i(y_i) + \sum_{y_i=c_i+1}^{m_{mi}} \alpha_i(c_i)\beta_i(y_i) +$$

$$\sum_{y_i=m_{mi}+1}^{2m_{mi}-1} \left[\alpha_i(c_i)\beta_i(m_i) - \sum_{j=0}^{y_i-m_{mi}-1} \frac{p(j)\rho_i}{c_i} \right].$$

进一步可得到各稳态概率*p_i(*y_i*)的值。*

为有效开展备件配置研究,必须掌握在补给周期时间*t* (0 ≤ *t* ≤ *T*)内SRU*i*备件的短缺情况。设备件短缺数为*n_i*,下面分不缺备件(*n_i* = 0)和备件短缺(*n_i* > 0)两种情况进行讨论^[11]。

当*n_i* = 0时,需满足条件:系统对SRU*i*备件的需求数*x_i* ≤ *S_{li}*,且基地级在修或待修SRU*i*的件数小于或等于*S_{0i}*。进而得到*T*时刻系统不缺SRU*i*备件的概率为:

$$\begin{aligned} Pr_i(BO_i = 0) &= P(S_{li} | m_i T) \cdot Pr_i(DI \leq S_{0i}) \\ &= \sum_{x_i=0}^{S_{li}} p(x_i | m_i T) \cdot \sum_{y_i=0}^{S_{0i}} p_i(y_i) \end{aligned} \tag{11}$$

当*n_i* > 0时,需满足条件:系统对SRU*i*备件的需求数*x_i* = *S_{li}* + *n_i*,且基地级在修或待修SRU*i*的件数小于或等于*S_{0i}*。进一步得到*T*时刻系统SRU*i*备件短缺数*n_i* (*n_i* ≥ 1)的概率为:

$$\begin{aligned} Pr_i(BO_i = n_i) &= p(S_{li} + n_i | m_i T) \cdot Pr_i(DI \leq S_{0i}) \\ &= p(S_{li} + n_i | m_i T) \cdot \sum_{y_i=0}^{S_{0i}} p_i(y_i) \end{aligned} \tag{12}$$

对于SRU*i*而言,每个LRU中有1个安装位置,则*N*个LRU中有*N*个安装位置。设*N*个LRU中SRU*i*备件短缺总数为*n_i*,若随机选取的1个LRU中SRU*i*的备件短缺数为0,则其余(*N* - 1)个安装位置中必定存在*n_i*个备件短缺数。根据超几何分布的知识可知,随机选取的任意1个LRU中SRU*i*备件短缺数为0的概率为:

$$p_i(n_i) = \frac{C_{N-1}^{n_i}}{C_N^{n_i}} = \frac{N-n_i}{N} \tag{13}$$

则 N 个 LRU 中未因 SRU*i* 备件短缺而故障的概率为:

$$p_i = \sum_{n_i=0}^{n_{i\max}} Pr_i(BO_i = n_i)p_i(n_i) \tag{14}$$

式中: $n_{i\max}$ 为 SRU*i* 备件短缺数的最大值, 满足 $n_{i\max} = m_i \circ$

从而可求得任意一个 LRU 正常工作的概率为:

$$p_0 = \prod_{i=1}^M p_i \tag{15}$$

根据 K/N 系统的定义, 进一步求得系统的供应可用度为:

$$A_s = \sum_{k=K}^N C_N^k p_0^k (1-p_0)^{N-k} \tag{16}$$

至此, 可以以备件配置费用最小为优化目标, 以系统供应可用度为约束条件构建系统的备件优化配置模型。

$$\begin{cases} \min & C \sum_{i=1}^M C_i (S_{0i} + S_{1i}) \\ \text{s. t.} & A_s \geq A_{s0} \end{cases} \tag{17}$$

式中: C_i 为第 SRU*i* 备件的单价; S_{0i} 为基地级 SRU*i* 备件的数目; S_{1i} 为基层级 SRU*i* 备件的数目; A_{s0} 为系统使用可接受的供应可用度最小值。

分析建立的优化模型可知, 模型的求解是一个大规模、非线性的非确定多项式 (Non-deterministic Polynomial, NP) 问题。在模型求解方法方面, 相对于遗传算法、粒子群优化算法等其他优化算法而言, 边际效应分析法操作简便、计算准确度高, 已成为国外一些先进的备件模型, 如瑞典的 OPUS10、美国的 VMETRIC 的核心算法^[12]。可见, 边际效应分析法已在实践中得到了检验, 是一种成熟的备件配置优化算法。为此, 本节将利用边际效应分析法对模型进行求解。

先求解备件满足率的边际效益值, 为此, 定义边际效益算子为:

$$\Delta S_{ji} = [A_s(S_{ji} + 1) - A_s(S_{ji})] / C_i \tag{18}$$

下面介绍具体算法。

Step 1: 确定系统控制变量, 即 S_{ji} 。令 $S = \{S_{01}, S_{02}, \cdots, S_{0M}, S_{11}, S_{12}, \cdots, S_{1M}\} = \{0, 0, \cdots, 0\}$, 记 S 中第 $h(1 \leq h \leq 2M)$ 个元素为 $S_h = S_{ji}$ 。

Step 2: 计算每一轮迭代过程中控制变量的最大边际效益值 $\Delta S_{h\max} = \max \{\Delta S_1, \Delta S_2, \cdots, \Delta S_{2M}\}$, 如果 $\Delta S_{h\max} = \Delta S_h$, 则 $S_h = S_h + 1$ 。

Step 3: 计算对应控制变量下的系统供应可用度的值 A_s , 如果 $A_s < A_{s0}$, 转到 Step 2; 否则, 算

法结束, 对应的控制变量 S 即为最终配置方案。

3 算例分析

某型 LPAR 天线阵面由 1440 个 T/R 组件 (LRU) 组成, 每个 T/R 组件主要由收发开关 (SRU1)、控制电路和保护电路 (SRU2)、功放 (SRU3)、移相器和限幅器 (SRU4) 等组成。整个天线阵面采用冗余设计, 可以看作一个 1220/1440 的 K/N 系统。天线阵面采用基层级和基地级两级维修体制, 并设有基层级库存和基地级库存, 其中, 基地级以固定时间 T 为周期定期补给基层级库存。

相控阵雷达故障诊断和维修设备的不断发展, 大大缩短了 LRU 换件维修的时间。因此, 只需在基地级和基层级配备足够多的 SRU*i* 备件即可满足系统的备件需求, 同时, 也可以大大减少备件的配置费用。已知: T/R 组件的寿命服从参数为 $\lambda = 0.000\ 5$ 的指数分布, 单个 LRU 的费用为 3 万元, LRU 换件维修人员 $c_r = 14$ 人, 换件维修率 $\mu_r = 2$, 各 SRU*i* 备件的具体参数见表 1。

表 1 SRU*i* 备件参数

Tab. 1 Spare parts parameters of SRU*i*

参数	SRU1	SRU2	SRU3	SRU4
λ_i	0.5×10^{-4}	1.5×10^{-4}	2×10^{-4}	1×10^{-4}
C_i /万元	1	0.6	0.5	0.8
c_i	3	6	8	5
μ_i	0.2	0.4	0.4	0.3

为满足天线阵面分系统供应可用度最低要求 $A_{s0} = 0.95$, 试根据提供的参数, 合理地对抗线阵面基层级和基地级的 SRU*i* 初始备件进行配置, 以使得备件总费用最小。

根据建立的维修资源优化配置模型, 代入相关数据, 运用模型求解算法对模型进行求解, 得到表 2 所示的优化结果。

表 2 SRU*i* 备件配置结果

Tab. 2 Configuration results of SRU*i* spare parts

	库存	备件 数目	供应 可用度	备件费 用/万元
基层级	S_{11}	18	0.957 2	196.6
	S_{12}	33		
	S_{13}	41		
	S_{14}	25		
	S_{01}	22		
基地级	S_{02}	66		
	S_{03}	43		
	S_{04}	44		

分析表2,基层级备件配置结果为 $S_{13} > S_{12} > S_{14} > S_{11}$,这与SRU故障率的大小 $\lambda_{SRU3} > \lambda_{SRU2} > \lambda_{SRU4} > \lambda_{SRU1}$ 是相符的,表明基层级备件的配置水平与SRU*i*的故障率是正相关的。基地级备件的配置结果则表现出 $S_{02} > S_{04} > S_{03} > S_{01}$,表明基地级备件配置水平受SRU*i*故障率 λ_i 、基地级维修渠道数 c_i 以及维修率 μ_i 等因素的多重影响。纵向来看,基地级的备件配置水平高于基层级,表明在现有基地级维修条件下,基地级备件数目需高于基层级备件数目才能满足系统的备件供应可用度要求。总的来说,本节提出的备件配置方法能够反映备件需求的基本规律,符合实际情况。

为了验证模型的有效性,下面通过比较分析的方法进行进一步的说明。

3.1 算例一

为了突出配置SRU备件的优越性,与利用单项备件配置方法配置LRU的情况进行了比较。利用单项备件配置方法对天线阵面的LRU备件进行配置,当 $n\lambda T > 5$ 时,备件需求量可利用正态分布近似计算,其计算公式^[1]为:

$$S = N\lambda T + u_p \sqrt{N\lambda T} \tag{19}$$

式中: u_p 为正态分布分位数,当要求保障概率为0.95时, $u_p = 1.65$ 。

根据上述计算方法,需配置LRU备件约267个,其备件费用为801万元。

综合比较可知:配置SRU备件比配置LRU备件可节约备件费用约604.4万元,降低约75.46%的经费。可见,配置SRU备件可以大大减少备件的配置费用,从而取得很好的经济和军事效益。

3.2 算例二

为了验证模型的有效性,保持其他各项参数不变,将维修能力增加一倍,即 $c_i = [6, 12, 16, 10]$ 。经过仿真,可得到维修能力增加后,各项备件的配置结果,见表3。

综合比较表2和表3可知:维修能力的提高,提高了基地级SRU*i*备件修复的效率,缩短了维修时间,从而打乱了原有SRU*i*备件配置之间的平衡关系,维修备件开始重新进行配置,备件总数大大减少,使得备件费用也相应降低近19.4%。然而,由于采用定期补给的方式,为保证基层级备件的需求,因此,出现基层级备件配置数目增加的情况,这与实际情况是相符合的。

图5给出了维修能力增加前后备件配置费用与系统可用度曲线的对比图。

表3 增加维修能力后的SRU*i*备件配置结果

Tab.3 Configuration results of SRU*i* spare parts after increasing repair capacity

	库存	备件数目	供应可用度	备件费用/万元
基层级	S_{11}	18	0.958 0	158.5
	S_{12}	43		
	S_{13}	67		
	S_{14}	27		
	S_{01}	22		
基地级	S_{02}	16		
	S_{03}	16		
	S_{04}	25		

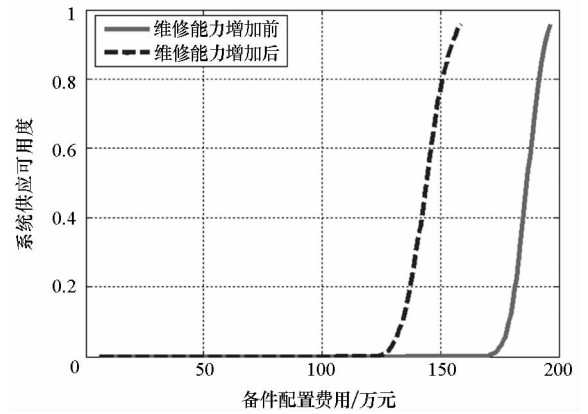


图5 维修能力增加前后备件配置对比图

Fig.5 Comparison chart of spare parts configuration before and after increase of repair capacity

从图5中可以看出:维修能力提高一倍后,花费很少的备件配置费用,系统供应可用度就可以达到 A_0 的最低要求;而当备件严重不足时,通过提高维修能力也可以得到较高的系统供应可用度。然而,在降低备件配置数目和备件配置费用的同时,增加系统的维修能力意味着维修费用也相应地增加。同时,从部队的角度来讲,编制体制也需相应变化,因此,一般不会通过增加维修能力的方式来降低备件配置水平。

4 结论

相控阵天线阵面的备件配置问题一直是部队亟待解决的难题。为此,本文从装备实际情况出发,综合考虑装备的冗余性、两级维修的特点以及维修能力的实际情况,提出了一种基于定期补给的相控阵天线阵面备件优化配置方法。构建了系统的供应可用度模型,以备件费用最小为目标对模型进行了优化,并运用边际效益分析法对模型进行了求解。算例仿真与分析结果表明:该方法

能够解决两级维修条件下考虑批量送修的大型冗余系统的维修备件配置问题,并且对于同一配置条件,相对于单独配置 LRU 的策略而言,该方法可以将维修备件的配置费用降低约 75.46%。此外,该方法能够结合装备与部队的实际情况对备件进行合理地配置,具有一定的应用价值。

参考文献 (References)

[1] 国防科学技术工业委员会. 备件供应规划要求: GJB 4355—2002[S]. 北京: 国防工业出版社, 2002.
Commission of Science and Technology Industry for National Defense. Spares provisioning requirements: GJB 4355—2002[S]. Beijing: National Defense Industry Press, 2002. (in Chinese)

[2] Yanagi S, Sasaki M, Umazume K. Optimal inventory problem of a repairable k -out-of- N : G system[J]. IEEE Transactions on Reliability, 1981, R-30(5): 478-480.

[3] 卢雷, 杨江平. $k/N(G)$ 系统初始备件配置方法[J]. 航空学报, 2014, 35(3): 773-779.
LU Lei, YANG Jiangping. Initial spare allocation method for $k/N(G)$ structure system [J]. Acta Aeronautica et Astronautica Sinica, 2014, 35(3): 773-779. (in Chinese)

[4] 薛陶, 冯蕴雯, 秦强. 考虑报废的 K/N 冷备份冗余系统可修复备件优化[J]. 华南理工大学学报: 自然科学版, 2014, 42(1): 41-46.
XUE Tao, FENG Yunwen, QIN Qiang. Optimization of repairable spare parts for K/N cold-standby redundant system considering scraps[J]. Journal of South China University of Technology: Natural Science Edition, 2014, 42(1): 41-46. (in Chinese)

[5] De Smidt-Destombes K S, Van Der Heijden M C, Van Herten A. On the interaction between maintenance, spare part inventories and repair capacity for a k -out-of- N system with wear-out [J]. European Journal of Operational Research, 2006, 174(1): 182-200.

[6] 张涛, 郭波, 武小悦, 等. k 阶段变化条件下 k/N : G 系统的备件保障度模型[J]. 兵工学报, 2006, 27(3): 485-488.

ZHANG Tao, GUO Bo, WU Xiaoyue, et al. Spare availability model for k -out-of- N system with different k in different phases[J]. Acta Armamentarii, 2006, 27(3): 485-488. (in Chinese)

[7] De Smidt-Destombes K S, Van Der Heijden M C, Van Herten A. Joint optimization of spare part inventory, maintenance frequency and repair capacity for k -out-of- N systems [J]. International Journal of Production Economics, 2009, 118(1): 260-268.

[8] 阮旻智, 李庆民, 彭英武, 等. 任意结构系统的备件满足率模型及优化方法[J]. 系统工程与电子技术, 2011, 33(8): 1799-1803.
RUAN Minzhi, LI Qingmin, PENG Yingwu, et al. Model of spare part fill rate for systems of various structures and optimization method [J]. Systems Engineering and Electronics, 2011, 33(8): 1799-1803. (in Chinese)

[9] 贾治宇, 王立超, 王乃超, 等. 基于停机时间的复杂系统维修资源配置模型[J]. 计算机集成制造系统, 2010, 16(10): 2211-2216.
JIA Zhiyu, WANG Lichao, WANG Naichao, et al. Maintenance resources configuration model for complex system based on down time[J]. Computer Integrated Manufacturing Systems, 2010, 16(10): 2211-2216. (in Chinese)

[10] 王乃超, 康锐. 备件需求产生、传播及解析算法研究[J]. 航空学报, 2008, 29(5): 1163-1167.
WANG Naichao, KANG Rui. Research on spare demand generation, transfer and analytical algorithm [J]. Acta Aeronautica et Astronautica Sinica, 2008, 29(5): 1163-1167. (in Chinese)

[11] Sherbrooke C C. Optimal inventory modeling of systems—multi-echelon techniques[M]. 2nd ed. Beijing: Publishing House of Electronics Industry, 2008.

[12] 阮旻智, 李庆民, 李承, 等. 改进的分层边际算法优化备件的初始配置方案[J]. 兵工学报, 2012, 33(10): 1251-1257.
RUAN Minzhi, LI Qingmin, LI Cheng, et al. Improved-layered-marginal algorithm to optimize initial spare part configuration project [J]. Acta Armamentarii, 2012, 33(10): 1251-1257. (in Chinese)

(上接第 171 页)

[11] 张晓峰, 胡庆波, 吕征宇. 基于 BUCK 变换器的无刷直流电机转矩脉动抑制方法[J]. 电工技术学报, 2005, 20(9): 72-76.
ZHANG Xiaofeng, HU Qingbo, LYU Zhengyu. Torque ripple reduction in brushless DC motor drives using a BUCK converter[J]. IEEE Transactions of China Electrotechnical Society, 2005, 20(9): 72-76. (in Chinese)

[12] Fang J C, Zhou X X, Liu G. Precise accelerated torque control for small inductance brushless DC motor[J]. IEEE Transactions on Power Electronics, 2013, 28(3): 1400-1412.

[13] Lin Y K, Lai Y S. Pulse width modulation technique for

BLDCM drives to reduce commutation torque ripple without calculation of commutation time[J]. IEEE Transactions on Industrial Application, 2011, 47(4): 1786-1793.

[14] Lu H F, Zhang L, Qu W L. A new torque control method for torque ripple minimization of BLDC motors with un-ideal back EMF[J]. IEEE Transactions on Power Electronics, 2008, 23(2): 950-958.

[15] Fang J C, Li H T, Han B C. Torque ripple reduction in BLDC torque motor with nonideal back EMF [J]. IEEE Transactions on Power Electronics, 2012, 27(11): 4630-4637.

加肋圆柱壳结构的 FE – IE 算法网格尺度划分原则*

黄振卫,周其斗,方 斌,谢剑波
(海军工程大学 舰船工程系,湖北 武汉 430033)

摘 要:为了研究结构有限元耦合流体无限元算法的网格尺度划分原则,提出有限元 – 无限元算法中结构湿表面的网格尺度划分原则。数值计算结果表明,对加肋圆柱壳而言,在保证一个肋骨间距至少有 2 个有限元单元的前提下,将结构主振型弯曲波波长作为有限元 – 无限元算法中结构湿表面网格尺度划分的参考标准是可行的,即保证一个主振型分量波长内至少有 6 个有限元单元。提出以结构主振型分量的弯曲波波长(而不是最短的结构波长)作为有限元 – 无限元算法中网格尺度划分的参考标准,所得结论对于有限元 – 无限元算法中结构湿表面的网格划分以及控制内域流体有限元数量均具有十分重要的参考意义。

关键词:加肋圆柱壳;有限元 – 无限元算法;振动与声辐射;网格尺度
中图分类号:U661.3 **文献标志码:**A **文章编号:**1001 – 2486(2017)03 – 179 – 06

Guidance of meshing scale of finite element coupled infinite element method for a ring stiffened cylindrical shell

HUANG Zhenwei, ZHOU Qidou, FANG Bin, XIE Jianbo

(Department of Naval Ship Engineering, Naval University of Engineering, Wuhan 430033, China)

Abstract: In order to investigate the meshing scale of the FE – IE (finite element coupled infinite element) method for a ring stiffened cylindrical shell, the guidance was put forward for the structural wetted surface. Numerical results show that, for ring-stiffened cylindrical shells, the bending wavelength of the main vibration mode can be taken as the reference of meshing scale in FE – IE calculation which is testified with at least two finite elements per distance of stiffeners. There are at least six finite elements per one bending wavelength of the main vibration mode. The bending wavelength of the main vibration mode is used as the reference of meshing scale of FE – IE calculation rather than the shortest wavelength of structural waves. The obtained conclusions are significant to the meshing of structural wetted surface and the controlling of the total number of elements in the interior fluid region.

Key words: ring stiffened cylindrical shell; finite element coupled infinite element method; vibration and sound radiation; mesh scale

有限元 – 无限元 (Finite Element coupled Infinite Element, FE – IE) 算法在声学中的应用研究比较活跃^[1-9], 新的理论和算法不断地出现和完善, 其中 Astley-Leis 无限元^[1] 和 Burnett 无限元^[2] 最具代表性。但是不管是哪种无限元, 其核心目标都是希望在保证计算精度的前提下尽量降低计算成本。对于一般结构而言, FE – IE 算法的人工边界的大小 r 可选取为结构几何直径的 3 倍^[3], 人工边界的单元尺度可选取为 1/6 倍的声波波长, 且靠近结构湿表面 (结构物与水接触的外表面) 的流体单元尺度等于结构湿表面上有限元单元尺度。此时, 为了节约计算成本, 应采用较大的单元尺度对结构湿表面进行有限元划分, 以便于减少内域流体的有限元数量, 但是这会由于单元尺度无法描述结构湿表面上以及近场流体中短波的波动特性而导致计算精度降低。另一方面, 为了保证计算精度, 结构湿表面的单元尺度以及近场流体的单元尺度应该尽量小, 但是这将导致实际大型复杂工程结构的求解规模急剧地增加, 使得 FE – IE 算法的计算时间非常长。因此对于 FE – IE 算法, 结构湿表面单元尺度的选取与计算精度和计算成本之间有个权衡的问题。目前很少有文献研究 FE – IE 算法中结构湿表面的单元尺度划分原则问题。

圆柱壳结构作为很多工程结构的典型结构^[10-11], 对其 FE – IE 算法的网格尺度划分进行研究具有一定的指导意义。结构的振动可以通过傅里叶变换分解为一系列不同波长的振动分量的

* 收稿日期:2016 – 01 – 25
基金项目:国家自然科学基金资助项目(51309230)
作者简介:黄振卫(1986—),男,湖北恩施人,博士研究生,E-mail: hzw125760220@126.com;
周其斗(通信作者),男,教授,博士,博士生导师,E-mail: qidou_zhou@126.com

叠加^[12],一般而言,应保证结构湿表面上最短的结构振动波长内至少有 6 个有限元单元^[13]。工程上,不能简单地采用最短的结构波长作为 FE-IE 算法中结构湿表面的有限元单元划分的参考标准,主要原因是通过分解得到的最短结构波长可以无限小,假如采用一个结构波长 6 个单元进行有限元划分,会导致结构湿表面以及内域流体的有限元数量呈级数式地增加,从而使得耦合系统的矩阵求解成本急剧增加。由于主振型分量的波长对流固耦合的计算具有重要的作用,因此文中以有限长加肋圆柱壳为研究对象,提出了 FE-IE 算法中结构湿表面的网格尺度划分原则,并通过 FE-IE 算法与有限元-边界元(Finite Element-Boundary Element, FE-BE)算法的数值计算结果的对比验证了该原则的有效性,FE-BE 算法程序的正确性在文献[14-15]中得到了很好的验证。

1 数学模型及有限元划分

以一个长 5.715 m,直径 1.27 m 的加肋圆柱壳为研究对象,壳体的内表面上均匀布置了 23 根肋骨,肋骨尺寸为 0.012 7 m×0.050 8 m,壳体厚度为 0.006 35 m,两端盖板厚度为 0.025 m。幅值为 4.454 N 的激励力作用于中间肋骨上,激励力的方向为沿着径向指向壳体外部。加肋圆柱壳的材料密度为 7850 kg/m³,泊松比为 0.3,弹性模量为 2.06×10¹¹ N/m²,结构阻尼为 0.06。假设模型置于无限水深且为自由边界条件,周围流体的密度为 1030 kg/m³。采用 PATRAN 软件进行有限元建模,壳体表面和两端的端盖采用三角形单元进行划分,肋骨采用四边形单元进行划分。采用 NASTRAN 软件中 IE 算法计算加肋圆柱壳的振动与声辐射。无限元的计算阶次选取为 10。IE 算法的人工边界选取为一个长 10 m,半径为 5 m(大于 3 倍的结构几何直径)的圆柱面。人工边界采用三角形单元进行划分,人工边界与结构湿表面之间的内域流体采用四面体单元进行划分,内域流体的单元尺度从结构湿表面到人工边界逐渐增大,靠近结构湿表面的流体单元尺度较小,靠近人工边界的流体单元尺度较大。用于比较的 72 个声场点均分布于与柱壳中间肋骨同心,半径为 9 m 的圆周上,如图 1 所示。由于加肋结构的湿表面最大网格尺度只能为 0.238 m,即一个肋骨间距一个单元,因此,为了便于研究,将计算频率选取为 100 Hz、200 Hz、300 Hz、500 Hz。

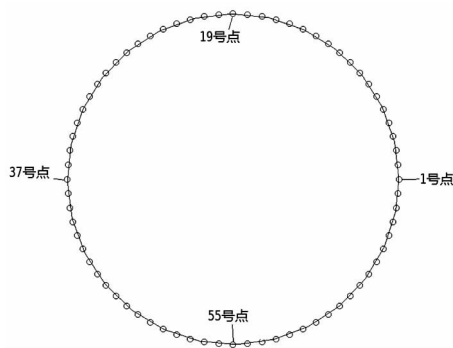


图 1 声场点分布

Fig. 1 Field point for acoustic computation

2 分离结构的主振型分量

本节采用自编波数谱程序分离有限长圆柱壳的主振型分量,该程序的正确性已在文献[16]中得到了很好的验证。为了描述有限长圆柱壳刚体运动的波数谱,将柱壳长度延长为原长度的 12 倍,并假定延长部分的外壳体法向速度为零^[16]。首先,采用 FE-BE 算法程序^[14]计算圆柱壳湿表面的法向位移复数幅值 $w(z, \theta)$, 即^[16]:

$$w(z, \theta) = \begin{cases} w_R + jw_I, & z \in [0, L] \\ 0, & z \notin [0, L] \end{cases} \quad (1)$$

式中, w_R 为结构湿表面的法向位移的实部, j 为虚数, w_I 为结构湿表面的法向位移的虚部, L 是圆柱壳的长度。为了保证提取的圆柱壳的湿表面法向位移的正确性,加肋圆柱壳的每档肋骨间距采用 6 个单元进行划分,周向采用 48 个单元进行划分。对法向位移的实部 w_R 沿柱壳的轴向和周向进行傅里叶变换,得到实部所代表的驻波场为:

$$\begin{aligned} w_R(z, \theta) e^{-j\omega\tau} &= \sum_{n=0}^{\infty} (a_n^R \cos n\theta + b_n^R \sin n\theta) e^{-j\omega\tau} \\ &= \frac{1}{2\pi} \sum_{n=0}^{\infty} \left\{ \int_{-\infty}^{+\infty} [A_n^R(k_z) \cos n\theta + \right. \\ &\quad \left. B_n^R(k_z) \sin n\theta] e^{-j\omega\tau} e^{jk_z z} dk_z \right\} \end{aligned} \quad (2)$$

式中, k_z 为轴向波数, n 为柱壳周向上完整波的数量, a_n^R 、 b_n^R 、 $A_n^R(k_z)$ 、 $B_n^R(k_z)$ 均为分解系数。采用相同的处理方法可获得虚部所代表的驻波场为:

$$\begin{aligned} w_I(z, \theta) e^{-j\omega\tau} &= \sum_{n=0}^{\infty} (a_n^I \cos n\theta + b_n^I \sin n\theta) e^{-j\omega\tau} \\ &= \frac{1}{2\pi} \sum_{n=0}^{\infty} \left\{ \int_{-\infty}^{+\infty} [A_n^I(k_z) \cos n\theta + \right. \\ &\quad \left. B_n^I(k_z) \sin n\theta] e^{-j\omega\tau} e^{jk_z z} dk_z \right\} \end{aligned} \quad (3)$$

圆柱壳法向速度的振动功率可表示为:

$$E_v^T = \sum_{n=0}^{\infty} \int_0^{+\infty} E_v(n, k_z) dk_z \quad (4)$$

式中, $E_v(n, k_z)$ 为振动分量 (n, k_z) 的法向速度振动功率谱,其表达式为:

$$E_v(n,k_z)=\begin{cases} \rho c w^2[|A_n^R(k_z)|^2+|A_n^I(k_z)|^2], & n=0 \\ \frac{\rho c w^2}{2}[|A_n^R(k_z)|^2+|A_n^I(k_z)|^2+|B_n^R(k_z)|^2+|B_n^I(k_z)|^2], & n\neq 0 \end{cases}$$

(5)

式中, a 为圆柱半径, ρ 为密度, c 为声速, w 为圆频率。

采用式(5)绘制出结构法向速度振动功率随着结构轴向波数的变化曲线,根据该曲线的最大值提取出结构主振型分量(n,k_z),进而获得结构主振型分量的振动波长 $\lambda_z=2\pi/k_z$ 。模型的法向速度振动功率谱($n=0,1,\cdots,9$)如图2所示,从图2(a)可以看出,当 $n=2,k_z\times 12L/\pi=34.01$ 时,结构在100 Hz时的主振型分量,对应的结构轴向波长为4.03 m。分析圆柱壳周向方向上结构波的传播特

性,可以得到 $n=2$ 时,周向上的结构波长为 $2\pi a/n=1.99$ m。采用相同的方法即可获得其他频率下结构的主振型分量,计算结果见表1。值得一提的是,图2中振动能量级大3 dB,意味着携带的能量将增大一倍左右。对于图2(a)~(c)而言,主振型分量显然高出其他振动分量3 dB以上。对于图2(d)而言,在主振型分量左边有两个其他振动模式下携带能量最多的振动分量(与主振型分量相比,振动能量要小2 dB左右),且这两个振动分量对应的结构弯曲波波长比主振型分量的弯曲波波长小,换言之,只要单元尺度能够描述主振型分量的波动特性,那么结构单元尺度便能描述这两个其他振动模式下携带能量最多的振动分量。因此,为了便于工程应用,500 Hz时仍然可以采用主振型分量的弯曲波波长作为单元划分的参考。

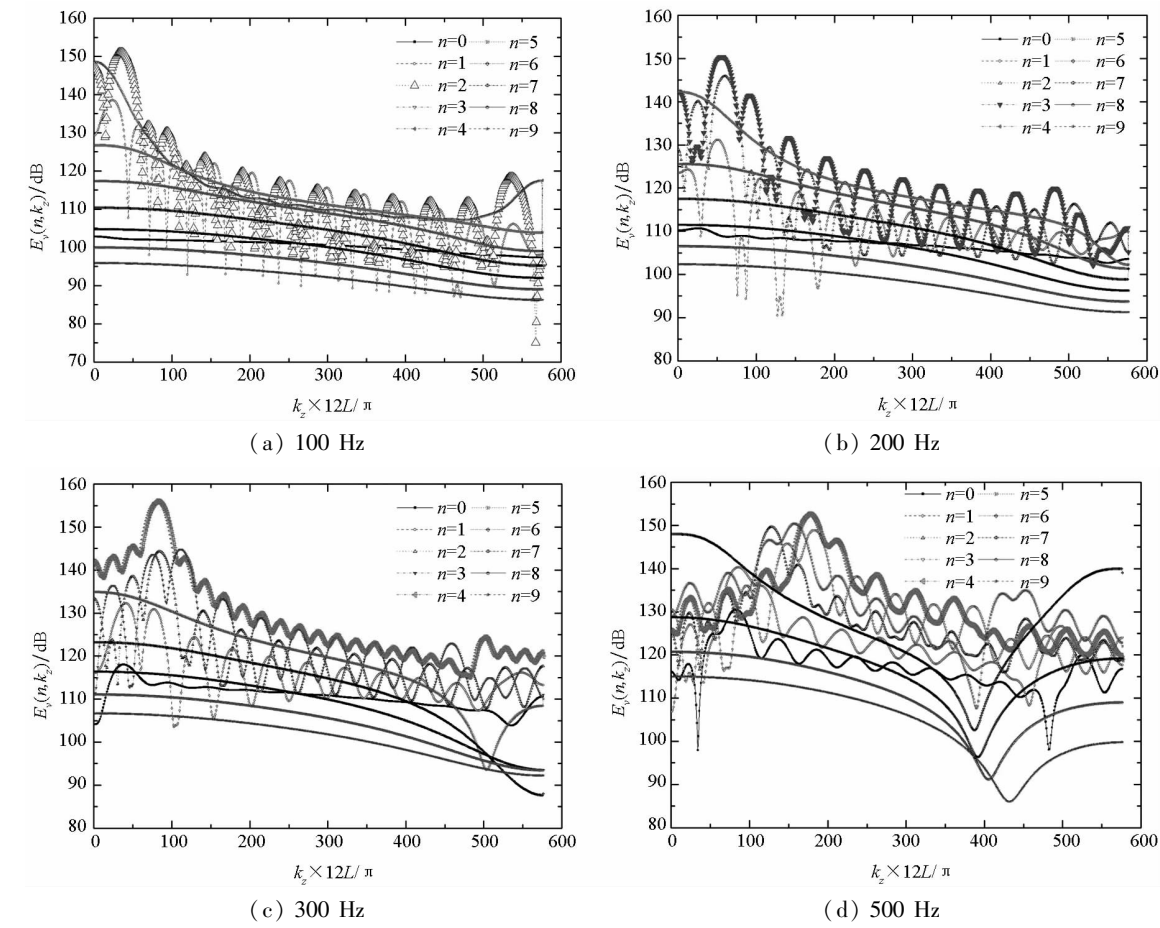


图2 模型法向速度振动功率谱

Fig.2 Normal velocity vibration power wavenumber spectrums of the model

表1 结构主振型分量

Tab.1 Main vibration mode

频率/Hz	n	λ_z/m	$(2\pi a/n)/\text{m}$	λ/m
100	2	4.03	1.99	14.5
200	3	2.45	1.33	7.25
300	4	1.67	1.00	4.83
500	4	0.77	1.00	2.9

图3为加肋圆柱壳模型的振型图,从图中可以看出,100 Hz时模型的周向振动模式主要为 $n=2$,200 Hz时模型的周向振动模式主要为 $n=3$,300 Hz、500 Hz模型的周向振动模式主要为 $n=4$,与波数谱分析结论一致;随着激励频率的增高,模型表面肋骨间的板格振动更加明显。

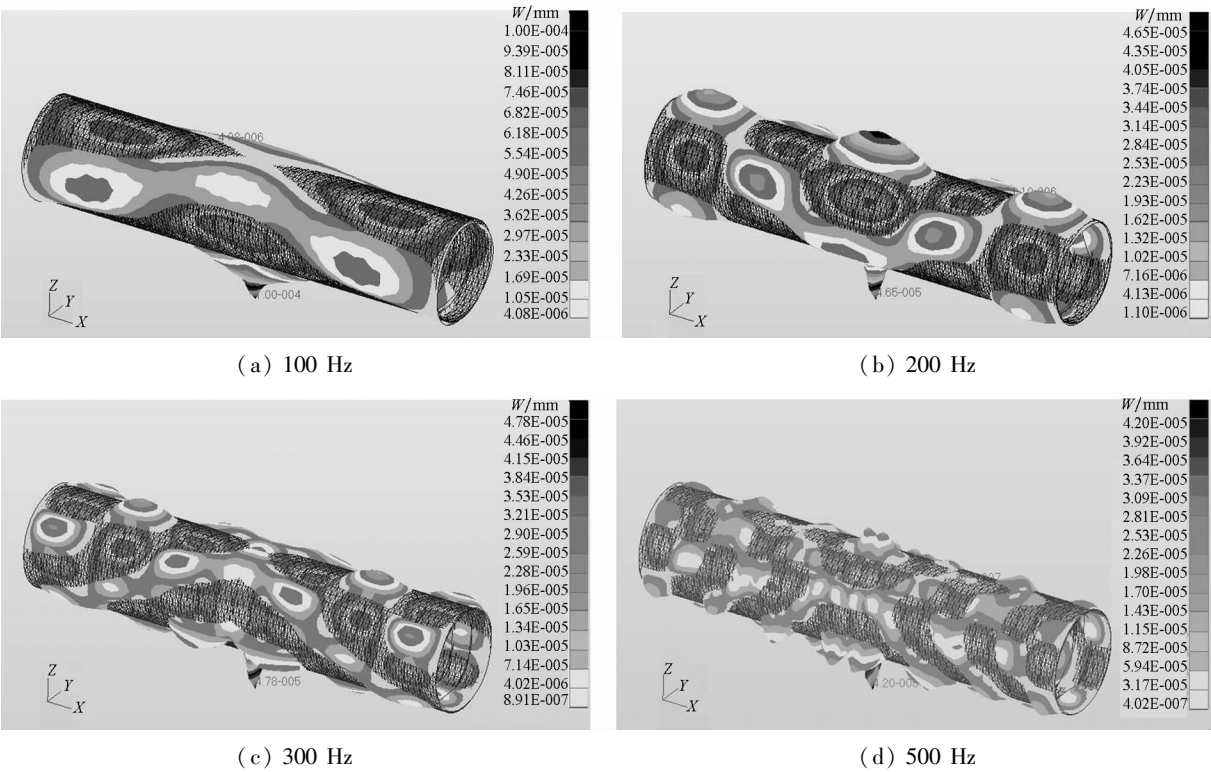


图 3 加肋圆柱壳模型的振型
Fig. 3 Vibration shape of the cylindrical shell mode

3 数值分析

由于加肋圆柱壳的振动与声辐射缺乏理论解,文中将逐步加密结构湿表面上的网格,观察 FE - IE 算法的计算结果是否趋于稳定,并与 FE - BE 程序^[14]计算结果进行对比。为了保证参与对照的 FE - BE 算法的计算精度,加肋圆柱壳的每档肋骨间距仍然采用 6 个有限元单元进行划分,周向采用 48 个有限元单元进行划分。根据第 2 节波数谱的计算结果,设计计算工况如表 2 所示,其中 L_z 为湿表面的有限元单元尺度, $N_z = \lambda_z / L_z$ 为一个主振型波长内的有限元单元数量, L_A 为人工边界的单元尺度。100 Hz 时, $L_A = 2.416$ m; 200 Hz 时, $L_A = 1.208$ m; 300 Hz 时, $L_A = 0.805$ m; 500 Hz 时, $L_A = 0.48$ m。

表 2 计算工况
Tab. 2 Computational cases

频率/Hz	N_z	L_z /m
100	101,68,34,17	0.04,0.059,0.119,0.238
200	61,42,21,10	0.04,0.059,0.119,0.238
300	42,28,14,7	0.04,0.059,0.119,0.238
500	19,13,6,3	0.04,0.059,0.119,0.238

图 4 ~ 7 为各工况下辐射声压级的 FE - IE 算法计算值对比,从图中可以看出:

1) 激励频率为 100 Hz 时,随着结构湿表面上有限元单元尺度的减小,FE - IE 算法的计算结果趋于稳定,且与 FE - BE 算法的计算结果吻合良好。结构湿表面上有限元单元尺度为 $L_z = 0.238$ m 时(一个肋骨间距一个有限元单元),虽然保证了一个主振型波长内 17 个有限元单元,但是 FE - IE 算法的计算结果与 FE - BE 算法的计算误差仍然较大;

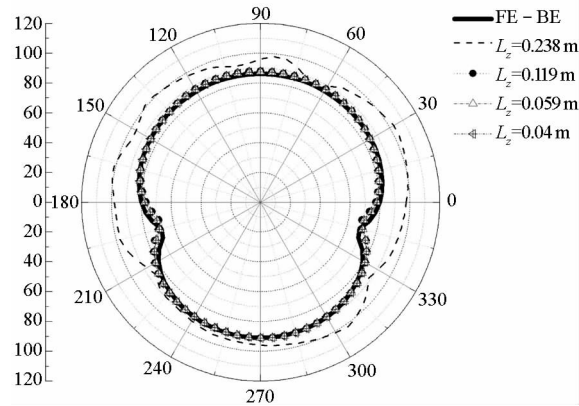


图 4 加肋圆柱壳在 100 Hz 时的声压级
Fig. 4 Numerical sound pressure level of the ring stiffened cylindrical shell at 100 Hz

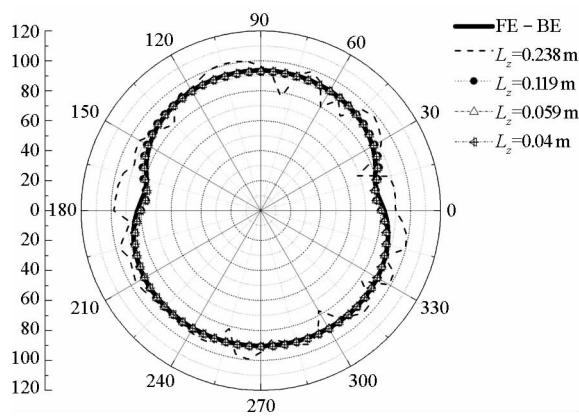


图5 加肋圆柱壳在 200 Hz 时的声压级
Fig.5 Numerical sound pressure level of the ring stiffened cylindrical shell at 200 Hz

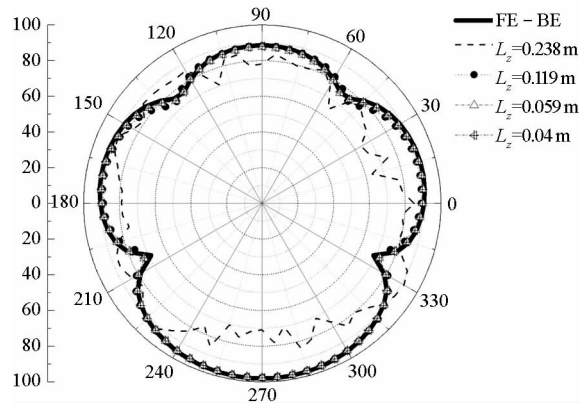


图6 加肋圆柱壳在 300 Hz 时的声压级
Fig.6 Numerical sound pressure level of the ring stiffened cylindrical shell at 300 Hz

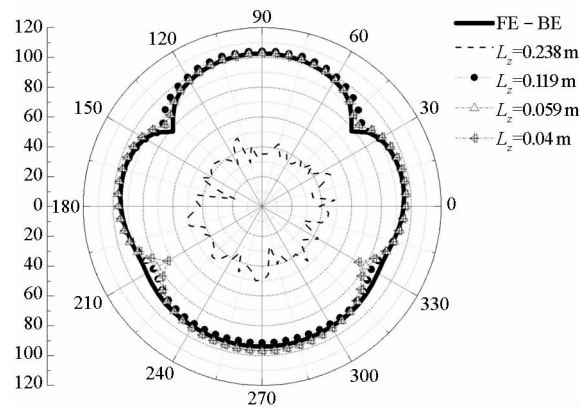


图7 加肋圆柱壳在 500 Hz 时的声压级
Fig.7 Numerical sound pressure level of the ring stiffened cylindrical shell at 500 Hz

2)激励频率为 200 Hz 时,随着结构湿表面单元尺度的减小,FE-IE 算法的计算结果趋于稳定,且与 FE-BE 算法的计算结果吻合良好。结构湿表面上有限元单元尺度为 $L_z = 0.238\text{ m}$ 时(一个肋骨间距一个有限元单元),虽然保证了一

个主振型波长内 10 个有限元单元,但是 FE-IE 算法的计算结果与 FE-BE 算法的计算误差仍然较大;

3)激励频率为 300 Hz 时,随着结构湿表面上有限元单元尺度的减小,FE-IE 算法的计算结果趋于稳定,且与 FE-BE 算法的计算结果吻合良好。结构湿表面上有限元单元尺度为 $L_z = 0.238\text{ m}$ 时(一个肋骨间距一个有限元单元),虽然保证了一个主振型波长内 7 个有限元单元,但是 FE-IE 算法的计算结果与 FE-BE 算法的计算误差仍然较大;

4)激励频率为 500 Hz 时,随着结构湿表面上有限元单元尺度的减小,FE-IE 算法的计算结果趋于稳定,但与 FE-BE 算法的计算结果有一定的差别。结构湿表面上有限元单元尺度为 $L_z = 0.238\text{ m}$ 时,一个主振型波长内只有大约 3 个有限元单元,此时 FE-IE 算法的计算结果与湿表面网格加密后的计算结果差别非常大。而结构湿表面上有限元单元尺度为 $L_z = 0.119\text{ m}$ 时,保证了一个肋骨间距 2 个有限元单元,一个主振型波长内大约有 6 个有限元单元,此时 FE-IE 算法的计算结果与湿表面网格加密后 FE-IE 算法的计算结果吻合良好。500 Hz 时湿表面网格加密后 FE-IE 算法的计算结果与 FE-BE 算法的计算结果有一定差别的原因是此时近场效应对 FE-IE 算法的影响更为明显,此时人工边界应该选取在距离结构湿表面更远的位置或者采用更小的有限元单元尺度划分人工边界。

总的来说,对于加肋结构而言,在保证一个肋骨间距至少 2 个有限元单元的前提下,将结构主振型弯曲波长作为 FE-IE 算法中结构湿表面网格尺度划分的参考标准是可行的,即保证一个主振型分量波长内至少有 6 个有限元单元。值得注意的是,此时结构湿表面上有限元单元尺度能够描述结构的几何特性和结构中携带能量最大的振动分量,进而保证了 FE-IE 算法的计算精度。若将结构湿表面上的有限元单元尺度进一步减小或者增大,都会导致对应的计算成本的增加或计算精度的损失。

为了进一步应用 FE-IE 算法的网格划分原则,本节将计算 50 ~ 500 Hz 时外域声场中的 $A(0\text{ m}, 0\text{ m}, 100\text{ m})$ 点的辐射声压级。为了保证计算精度,加肋圆柱壳每个肋骨间距采用 2 个有限元单元进行划分($L_z = 0.119\text{ m}$),人工边界为一个长 10 m,半径为 5 m 的圆柱面,人工边界的单元尺度为 0.48 m(500 Hz 时保证一个声波波长内

有 6 个单元)。A 点的辐射声压级传递函数频响曲线如图 8 所示,从图中可以看出,激励频率较低时,FE-IE 算法的辐射声压级传递函数频响曲线与 FE-BE 算法的频响曲线吻合良好;激励频率较高时,FE-IE 算法的辐射声压级传递函数频响曲线与 FE-BE 算法的频响曲线变化趋势一致。

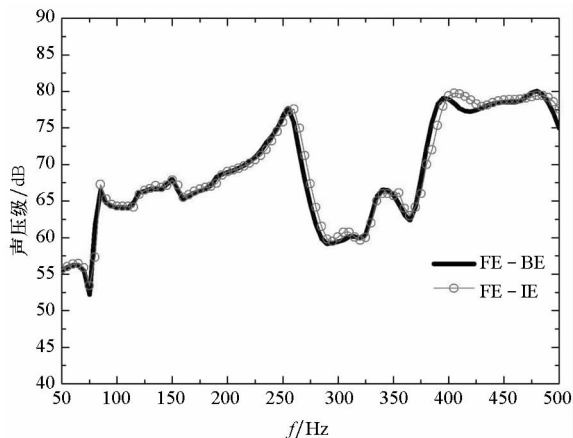


图 8 结构声学传递函数的 FE-IE 计算值

Fig. 8 Sound transfer function obtained by the FE-IE method

4 结论

本文以加肋圆柱壳为研究对象,重点采用波数谱方法分离结构的主振型分量,根据主振型分量的弯曲波波长提出 FE-IE 算法中结构湿表面的网格尺度划分原则,并采用 FE-IE 算法验证了该原则的有效性,计算结果表明,对文中算例而言,在保证一个肋骨间距至少 2 个有限元单元的前提下,将结构主振型弯曲波长作为 FE-IE 算法中结构湿表面网格尺度划分的参考标准是可行的,即保证一个主振型分量波长内至少有 6 个有限元单元。所得结论对于 FE-IE 算法的有限元模型的网格划分以及控制内域流体有限元数量具有十分重要的参考意义。

参考文献 (References)

- [1] Astley R J, Hamilton J A. Numerical studies of conjugated infinite elements for acoustical radiation [J]. *Journal of Computational Acoustics*, 2000, 8(1): 1-24.
- [2] Burnett D S. A three-dimensional acoustic infinite element based on a prolate spheroidal multipole expansion [J]. *The Journal of the Acoustical Society of America*, 1994, 96(5): 2798-2816.
- [3] Astley R J, Eversman W. Wave envelope and infinite element schemes for fan noise radiation from turbofan inlets [J]. *AIAA Journal*, 2012, 22(12): 1719-1726.
- [4] Burnett D S, Holford R L. An ellipsoidal acoustic infinite element [J]. *Computer Methods in Applied Mechanics and Engineering*, 1998, 164(1/2): 49-76.
- [5] Shirron J J, Babuska I. A comparison of approximate boundary conditions and infinite element methods for exterior Helmholtz problems [J]. *Computer Methods in Applied Mechanics and Engineering*, 1998, 164(1/2): 121-139.
- [6] Cremers L, Fyfe K R, Coyette J P. A variable order infinite acoustic wave envelope element [J]. *Journal of Sound and Vibration*, 1994, 171(4): 483-508.
- [7] Dreyer D, Petersen S, VonEstorff O. On the efficiency of exterior acoustics simulations using improved infinite elements [C]//*Proceedings of the International Conference on Noise and Vibration Engineering, ISMA, 2004*: 3778-3791.
- [8] 吴国荣. 声辐射问题的一种新的无限元方法[J]. *船舶力学*, 2009, 13(4): 641-645.
WU Guorong. A new infinite element method for acoustical radiation problems [J]. *Journal of Ship Mechanics*, 2009, 13(4): 641-645. (in Chinese)
- [9] 杨瑞梁, 范晓伟. 使用有限元和无限元耦合求解声辐射问题[J]. *振动工程学报*, 2004, 17(2): 1007-1009.
YANG Ruiliang, FAN Xiaowei. Finite/infinite element method for the acoustic radiating problem [J]. *Journal of Vibration Engineering*, 2004, 17(2): 1007-1009. (in Chinese)
- [10] 梁波. 弹性储液圆柱壳的动力特性分析[J]. *国防科技大学学报*, 1990, 12(2): 30-35.
LIANG Bo. The dynamic characteristics analysis of flexible liquid storage cylinder [J]. *Journal of National University of Defense Technology*, 1990, 12(2): 30-35. (in Chinese)
- [11] 汤渭霖, 何兵蓉. 水中有限长加肋圆柱壳体振动与声辐射近似解析解[J]. *声学学报*, 2001, 26(1): 1-5.
TANG Weilin, HE Bingrong. Approximate analytic solution of vibration and sound radiation from stiffened finite cylindrical shells in water [J]. *Acta Acustica*, 2001, 26(1): 1-5. (in Chinese)
- [12] Fahy F J, Gardonio P. Sound and structural vibration [M]. Academic Press, Oxford: Elsevier, 2007.
- [13] 宗福开. 波传播问题中有限元分析的频散特性及离散化准则[J]. *爆炸与冲击*, 1984(4): 18-25.
ZONG Fukai. Frequency dispersion characteristic and discretization of the finite element analysis in wave propagation problems [J]. *Explosion and Shock Waves*, 1984(4): 18-25. (in Chinese)
- [14] Zhou Q, Joseph P F. A numerical method for the calculation of dynamic response and acoustic radiation from an underwater structure [J]. *Journal of Sound and Vibration*, 2005, 283(3/4/5): 853-873.
- [15] Zhou Q, Zhang W, Joseph P F. A new method for determining acoustic added mass and damping coefficients of fluid-structure interaction [C]//*Proceedings of the 8th International Symposium on Practical Design of Ships and Other Floating Structures*, 2001: 1185-1195.
- [16] 谭路, 纪刚, 张纬康, 等. 采用波数域方法分析细长柱壳的振动与声辐射特性[J]. *海军工程大学学报*, 2013, 25(3): 66-71.
TAN Lu, JI Gang, ZHANG Weikang, et al. Slender cylindrical vibration and radiation by use of wave-number domain approach [J]. *Journal of Naval University of Engineering*, 2013, 25(3): 66-71. (in Chinese)

纤维织物复合材料组分材料体分比的显微CT实验测定法*

王浩,王中伟

(国防科技大学 高超声速冲压发动机技术重点实验室, 湖南 长沙 410073)

摘要:针对纤维织物复合材料的组分材料体分比测定问题,提出一种基于显微CT图像的测定方法。该方法可以通过不同尺度的显微CT图像分别测定全局纤维体分比、局部纤维体分比和纤维束体分比参数,还可以为难以用常规物理实验测定体分比的复合材料组分材料体积分数测定提供解决方案。以E-Glass/Epoxy纤维织物复合材料为研究对象,对比ASTM D3171 Procedure G、扫描电镜实验和显微CT实验三种测定法的测量值,结果证明了显微CT实验测定法的可行性和合理性。针对扫描电镜图像和显微CT图像,分别给出了相应的图像处理方法,为获得正确的组分材料分割结果提供了技术保证。显微CT实验测定方法可以广泛应用于复合材料组分材料体分比的测定。

关键词:玻璃纤维/环氧树脂;纤维织物复合材料;体分比;显微CT;扫描电镜;图像分割

中图分类号:TB332 **文献标志码:**A **文章编号:**1001-2486(2017)03-185-09

Volume fraction measurement for component material of textile composite using micro CT experiments

WANG Hao, WANG Zhongwei

(Science and Technology on Scramjet Laboratory, National University of Defense Technology, Changsha 410073, China)

Abstract: A method for measuring the volume fractions of component material of textile composite by using micro CT experiments was developed. This method can present global, local fiber and yarn volume fractions by micro CT images in different scales, and can also offer solutions to the difficult volume fractions measurement of component materials of some composites which cannot be measured directly by conventional physical experiments. An E-Glass/Epoxy textile composite was used to illustrate the feasibility and reasonability of the method by the comparisons of the measured values among ASTM D3171 Procedure G, scanning electron microscope and micro CT experiments. Corresponding image processing methods for the scanning electron microscope and micro CT images were used to acquire the accurate component material segmentations. The measurement of micro CT experiments can be widely applied to measure the volume fractions of component materials of composite.

Key words: E-Glass/Epoxy; textile composite; volume fraction; micro CT; scanning electron microscope; image segmentation

近年来,显微CT(micro computed tomography)技术已经成为研究复合材料微观结构的一种重要手段。Desplentere等^[1]通过比较4种不同3D编织复合材料的纤维束厚度、宽度及间距参数在表面扫描照片、光学显微照片和显微CT图像中的测量值,说明了编织复合材料显微CT图像反映微观结构的可靠性。Madra等^[2]利用显微CT技术对织物复合材料中的孔隙分布进行了多尺度分析。Pazmino等^[3]利用显微CT技术研究了面内剪切变形对三维正交编织复合材料纤维束几何参数的影响。Schell等^[4]利用显微CT技术对纤维织物复合材料中的纤维束几何形状和孔隙进行了量化研究。一种基于显微CT图像的典型统计方法用于研究织物复合材料纤维束的微观结构特

征^[5-6]。Wang等^[7]利用高分辨率显微CT技术研究了C/Epoxy织物复合材料纤维束特征参数的统计特征。

在织物复合材料的力学计算模型^[8-12]中,纤维束在复合材料中的体积分数(简称纤维束体分比) V_y 和纤维在纤维束中的体积分数(简称局部纤维体分比) V_f 是两个重要的参数。这两个参数虽然可以通过理论计算^[13-16]近似得到,但真实建模中需要精确测量这两个参数来消除误差。通常情况下,通过物理实验,如ASTM D3171 Procedure G^[17](简称D3171 G实验)或化学消蚀法,只能测得纤维在材料中的体积分数(简称全局纤维体分比) \widehat{V}_f ,无法直接测得局部纤维体分比 V_f 和纤维

* 收稿日期:2016-01-18

作者简介:王浩(1984—),男,湖南常德人,博士研究生,E-Mail:gfkdw@163.com;

王中伟(通信作者),男,教授,博士,博士生导师,E-Mail:gfkdwz@163.com

束体分比 V_y 。此外,C/SiC 等陶瓷基复合材料无法通过 D3171 G 实验获得组分材料体分比。

显微 CT 技术是一种可以获得组分材料体分比的新方法。本文针对 E-Glass/Epoxy 纤维织物复合材料分别通过三种实验测定方法(D3171 G 实验、扫描电镜(Scanning Electron Microscope, SEM)实验和显微 CT 实验)测定了其组分材料体分比,并比较了三种实验测定方法的结果,说明了显微 CT 实验测定复合材料组分材料体分比的可行性和合理性。

纤维织物复合材料组分材料体分比的显微 CT 实验测定法可以通过不同尺度的显微 CT 图像测定全局、局部纤维体分比和纤维束体分比。此外,该方法还可以为难以用常规物理实验测定体分比的复合材料(如 C/SiC)组分材料体积分数测定提供解决方案。

1 材料与样品

1.1 材料制备

E-Glass/Epoxy 纤维织物复合材料为由 15 层平纹布叠加而成的层合板,每个铺层由一个名义单胞尺寸为 4 mm × 4 mm 的 E-Glass 平纹布组成。材料基体是由 Dow Chemical Company 提供的一类热固性环氧树脂 Airstone 760E 和固化剂 Airstone 766H 按重量以 100 : 32 的比例混合而成。该层合板是在一个预热的平板模具上通过真空导入工艺加热至 70 ℃ 并恒温 7 h 制备而成,然后自然冷却至室温。

1.2 实验样品的制备

1)D3171 G 实验样品的制备。将层合板切割成 25 mm × 25 mm 的方形样品共 3 个,并对其进行超声清洗。

2)扫描电镜实验样品的制备。为了在样品截面上获得纤维的清晰扫描电镜像,需要对样品截面进行以下三步操作:①依次使用 800 目、2000 目和 10 000 目的金刚石研磨膏对样品完成“打磨—抛光”操作;②对已完成抛光的样品进行超声清洗,去除样品表面的杂质和残留的金刚石研磨膏;③对样品的观察面(即抛光面)进行喷铭。

3)显微 CT 实验样品的制备。根据显微 CT 成像原理(见 4.2 节),要获得不同尺度下的显微 CT 图像,需要不同尺寸的显微 CT 实验样品。E-Glass/Epoxy 纤维织物复合材料的显微 CT 中尺度和微尺度成像样品尺寸分别为 20.20 mm × 6.58 mm × 3.84 mm,20.20 mm × 3.50 mm ×

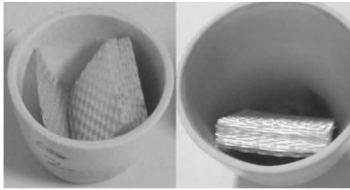
3.50 mm。样品是通过“高压水切割→手工打磨→超声清洗”的方式得到的。

2 全局纤维体分比的 D3171 G 实验测定

ASTM D3171 Procedure G 实验只能测定 E-Glass/Epoxy纤维织物复合材料的全局纤维体分比 \widehat{V}_f 。将制备的 3 个样品放入马弗炉中升温至 600 ℃ 并恒温 6 h 至样品中的环氧树脂基体完全烧尽。燃烧前后的样品如图 1 所示。利用分析天平对燃烧前后的 3 个样品进行称重并计算全局纤维体分比 \widehat{V}_f ,如表 1 所示。经测定,E-Glass/Epoxy 纤维织物复合材料的全局纤维体分比 \widehat{V}_f 的均值为 0.510 0,标准差为 0.003 4。



(a) 实验前
(a) Before experiment



(b) 实验后
(b) After experiment

图 1 E-Glass/Epoxy 纤维织物复合材料的 D3171 G 实验
Fig. 1 D3171 G experiment of E-Glass/Epoxy textile composite

表 1 全局纤维体分比 \widehat{V}_f 的 D3171 G 实验测定结果
Tab. 1 Global fiber volume fraction \widehat{V}_f measured by D3171 G experiment

样品 1	样品 2	样品 3	均值	标准差
0.506 1	0.511 3	0.512 5	0.510 0	0.003 4

3 局部纤维体分比的扫描电镜实验测定

纤维织物复合材料的局部纤维体分比 V_f 可以通过对纤维束截面的扫描电镜像进行统计分析来近似获得。但是,全局纤维体分比 \widehat{V}_f 和纤维束体分比 V_y 无法通过扫描电镜实验测定。由于局

部纤维体分比 V_f 随纤维束位置的变化而变化,故通过纤维束扫描电镜图像测定的局部纤维体分比 V_f 只能作为其近似估计值。E-Glass/Epoxy 纤维织物复合材料的扫描电镜像是通过双束扫描电镜

Helios Nanolab 600i 来得到的。选取样品表面的 3 个任意法向纤维束区域进行成像,如图 2 所示。对这 3 个纤维束区域进行局部纤维体分比 V_f 的测定和统计分析。

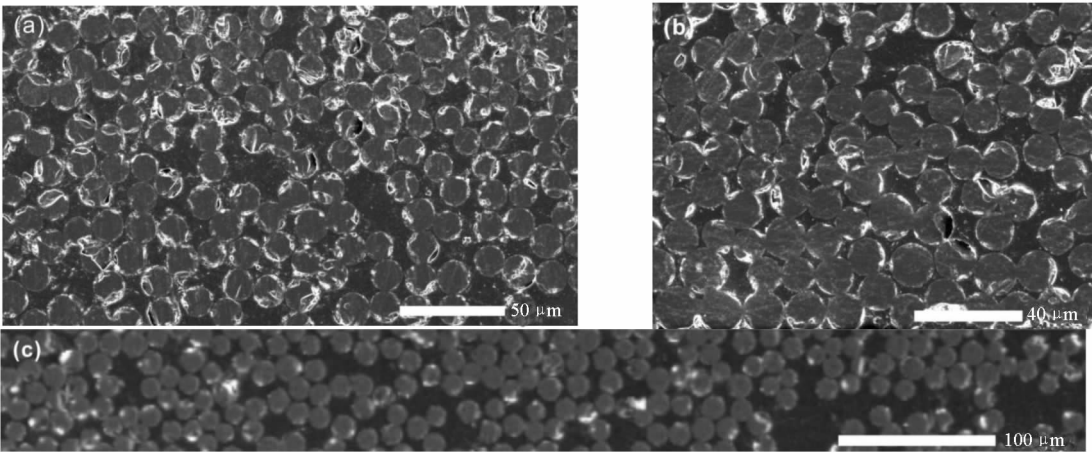


图 2 E-Glass/Epoxy 纤维织物复合材料的 3 个任意法向纤维束区域 SEM 图像
Fig. 2 SEM images of 3 arbitrary normal yarn areas of E-Glass/Epoxy textile composite

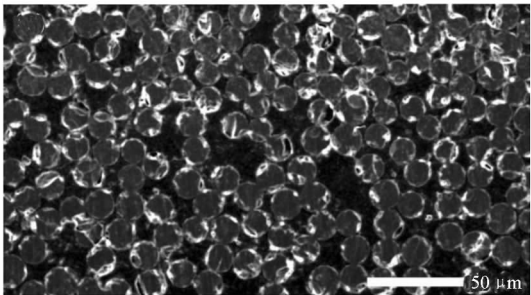
在组分材料分割前需要对该图像进行预处理以进一步提升图像中纤维和基体间的灰度值差异。该预处理过程在 Fiji 软件中完成,共分为四步:

- 1) 去除椒盐噪声 (Despeckle 功能);
- 2) 抑制纤维边缘明亮区域 (Remove Outlier 功能);
- 3) 对图像进行中值过滤 (Median 功能) 以模糊纤维或基体内部细节并保留纤维边缘;
- 4) 提升图像中纤维和基体间的对比度 (Enhance Contrast 功能)。

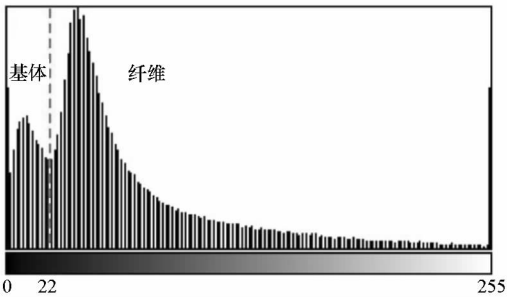
图 3(a)是对图 2(a)进行预处理后的纤维束区域图像。从图中可以明显看出纤维和基体内部的细节进行了均匀化处理,但纤维与基体间的边界仍然可以清晰辨认。图 3(a)中的明亮区域较图 2(a)要更加明显,这是对图像进行了提升对比度操作而成的。图 3(b)是图 3(a)的灰度直方图。从图中可以看出基体与纤维间的灰度值差异较原图有明显改善,在两组分材料间出现了较明显的灰度值谷点。

对预处理后的图像直接采用灰度阈值或分水岭分割算法无法得到满意的分割结果,这是因为在基体区域仍然存在较多的“杂质”区域。这些“杂质”区域的灰度值处于纤维和基体之间,经过一系列上述预处理操作后仍难以完全消除。

为了解决该问题,采用 Fiji 软件中可训练的 Weka 分割算法 (Trainable Weka Segmentation 插件)对图 3(a)进行组分材料区域分割。可训练



(a) 经预处理的法向纤维束区域图像
(a) Image of normal yarn area after preprocessing



(b) 对应的灰度直方图
(b) Corresponding gray level histogram

图 3 扫描电镜图像的预处理结果
Fig. 3 Preprocessing result of SEM image

的 Weka 分割算法通过对一系列选取的图像特征结合机器学习算法来产生基于像素的分割结果。

图 4 是对图 3(a)进行 Weka 分割后的二值化图像。从图 4 中可以看出,经过训练和反复改进,Weka 分割算法对纤维和基体进行了较为成功的分割,避开了基体中“杂质”图像的干扰。通过

计算图 4 所示两种材料区域(黑色为纤维,白色背景为基体)的面积可以测得图 2(a)对应的局部纤维体分比 $V_f=0.750$ 。

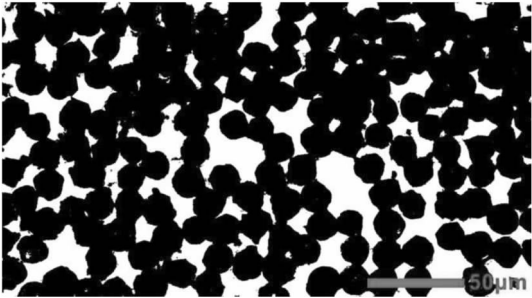


图 4 法向纤维束区域图像的 Weka 分割结果
Fig.4 Weka segmentation result of the normal yarn area image

类似地,用上述图像处理方法对图 2(b)和图 2(c)进行纤维和基体区域分割。通过测量两种材料区域的面积,可以测得对应的局部纤维体分比 V_f 分别为 0.840 和 0.656。三个纤维束区域 V_f 测定值的均值和标准差如表 2 所示。

表 2 局部纤维体分比 V_f 的扫描电镜实验测定结果				
Tab.2 Fiber volume fraction V_f measured by SEM experiment				
图 2(a)	图 2(b)	图 2(c)	均值	标准差
0.750	0.840	0.656	0.749	0.075

从表 2 可以看出,三个区域测定的局部纤维体分比 V_f 变化较大,这是由于扫描电镜测定结果与以下四个因素密切相关:

1) 所观测纤维束区域在纤维束中的位置;

- 2) 所观测纤维束区域是否与纤维束路径垂直;
- 3) 所测定纤维束区域图像的尺度大小;
- 4) 图像处理造成的误差。

其中因素 1~3 是影响扫描电镜测定结果的主要因素。此外,扫描电镜测定法还有一个隐含的假定前提:纤维沿纤维束轴线方向其截面面积不变,即该测定法是用纤维与基体所占面积的比值来得到局部纤维体分比,而非体积比。

4 组分材料体分比的显微 CT 实验测定

纤维织物复合材料的组分材料体分比(全局纤维体分比 \bar{V}_f 、局部纤维体分比 V_f 和纤维束体分比 V_y)可以通过对材料不同尺度的显微 CT 图像进行分析获得。

4.1 纤维织物复合材料的尺度划分

纤维织物复合材料是一种多级材料,可以划分为三个尺度^[18],如图 5 所示。将纤维织物复合材料的结构件(如飞机结构件、桥梁结构件等)列入宏观尺度,该尺度在米至千米级。

将层合板分解成单个铺层,每个铺层可以分解为纤维束类属和基体类属。该层级为中尺度(毫米级),也称为纤维束尺度。在该层级中将纤维束视为横观各向同性固体,不区分纤维束内部的纤维丝和基体。

将每根纤维束再分解成纤维和基体。该层级为微尺度(微米级),也称为纤维尺度。

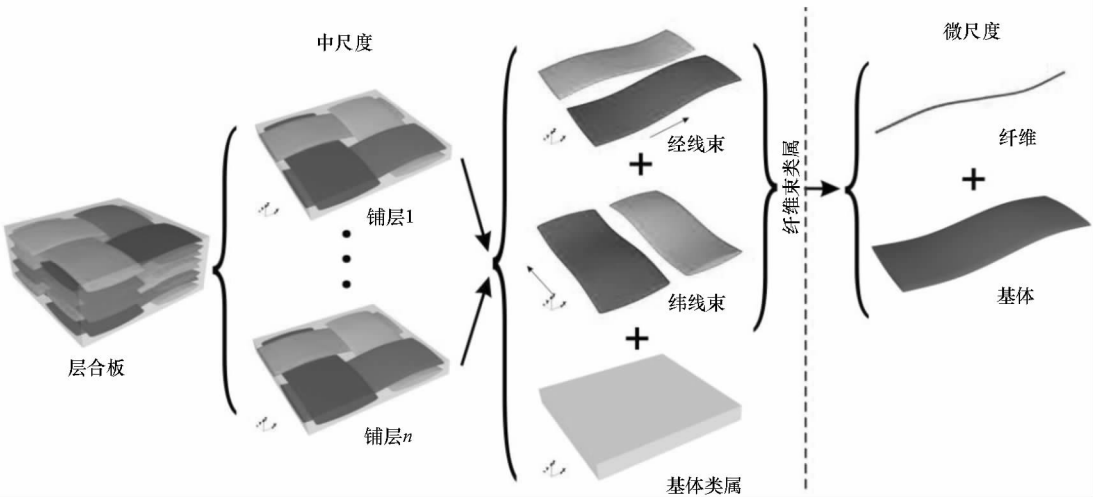


图 5 平纹编织复合材料的尺度划分
Fig.5 Scales of plain-weave composite

4.2 显微CT成像原理

显微CT是一种成像方法,用于重构目标对象的内部结构。一个显微CT系统通常由微焦点X射线源、样品台和高分辨率X射线探测器系统构成。有两类典型探测器系统用于显微CT系统:X射线平板探测器和高分辨率光耦探测器。

4.2.1 基于平板探测器的显微CT成像原理

X射线平板探测器用于大多数显微CT设备,它可以获得宽视场,如NIKON XTH225 ST。这类显微CT设备的放大原理为投影放大(即几何放大),如图6所示。基于平板探测器的显微CT成像原理为吸收衬度成像。这类设备对于高

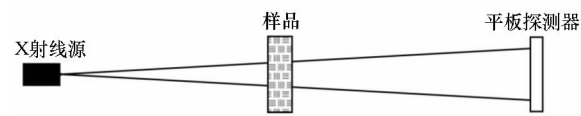


图6 基于平板探测器的显微CT成像原理图
Fig.6 Imaging principle of micro CT with flat panel detector

原子序数和高密度材料(如E-Glass/Epoxy等)可以获得清晰的影像。通过这类设备对纤维织物复合材料进行三维成像可以获得中尺度显微CT图像。

4.2.2 基于光耦探测器的显微CT成像原理

Sanying nanoVoxel-2000拥有一个高分辨率光耦探测器,其成像原理如图7所示。该设备有两级光学放大^[19]。第一级光学放大是几何放大,类似于平板探测器的放大原理。第二级放大是通过一系列光学物镜实现的。两级光学放大的优点是无需将样品靠近X射线光源就可以实现高放大倍率。高分辨率光耦探测器的最终成像设备是一个冷却的科学级CCD相机。该相机可以探测非常微弱的光信号。此外,该设备利用相位衬度来进行成像,可以获得更高的图像灵敏度。这些技术为获取复合材料的高分辨率图像提供了技术保障。通过这类设备对纤维织物复合材料进行三维成像可以获得中尺度或微尺度显微CT图像。

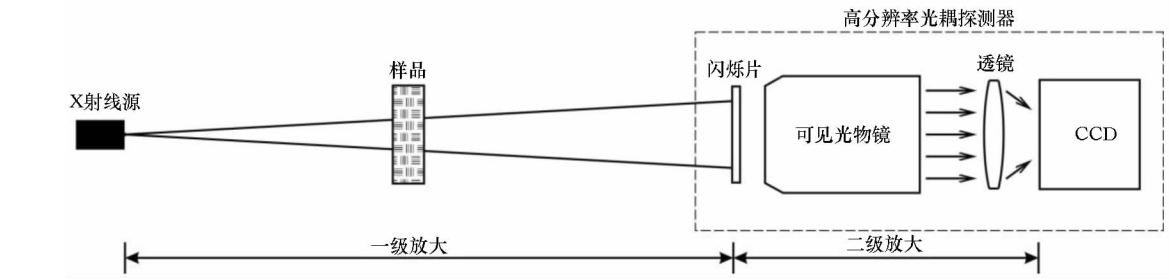


图7 基于光耦探测器的显微CT成像原理图
Fig.7 Imaging principle of micro CT with high-resolution lens-coupled detector

4.3 多尺度显微CT图像的获取与处理

4.3.1 中尺度显微CT图像的获取

E-Glass/Epoxy纤维织物复合材料的中尺度图像是通过具有平板探测器的NIKON XTH225 ST扫描获得的。该设备拥有一个180 kV/1 mA的微焦点射线管,最高分辨率为1 μm 。对于中尺度成像样品,在145 kV/124 μA 条件下获得的图像分辨率为6.129 μm ,如图8所示,图像区域大小为10.149 mm \times 5.651 mm \times 3.843 mm。通过中尺度显微CT图像可以分析中尺度下的孔隙、裂纹以及纤维束在复合材料中的体积含量等。

4.3.2 微尺度显微CT图像的获取

E-Glass/Epoxy纤维织物复合材料的微尺度图像是通过具有光耦探测器的Sanying nanoVoxel-2000扫描获得的。该设备拥有一个150 kV/0.5 mA的微焦点X射线源,最高分辨率为500 nm。对于微尺度成像样品,在80 kV/100 μA 的参数下获得分辨率为1.036 μm 的图像,如图9

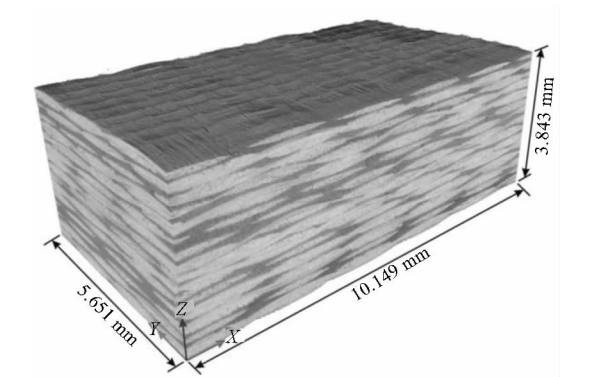


图8 E-Glass/Epoxy纤维织物复合材料的中尺度三维显微CT图像
Fig.8 3D micro CT image of E-Glass/Epoxy textile composite in meso-scale

所示,图像区域大小为0.947 mm \times 0.873 mm \times 0.575 mm。通过微尺度显微CT图像可以分析微尺度下的孔隙、裂纹以及纤维在纤维束中的体积含量和全局纤维体分比等。

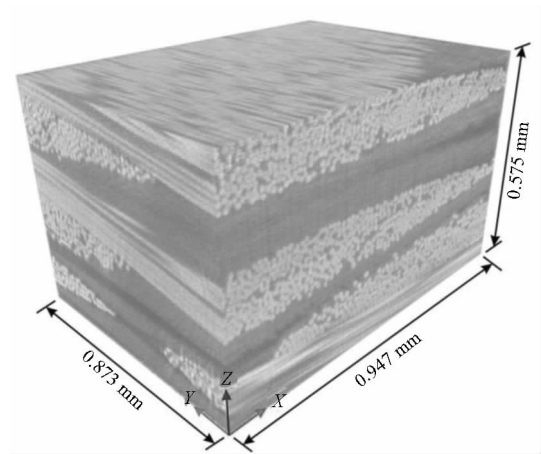


图9 E-Glass/Epoxy 纤维织物复合材料的微尺度三维显微 CT 图像

Fig. 9 3D micro CT image of E-Glass/Epoxy textile composite in micro-scale

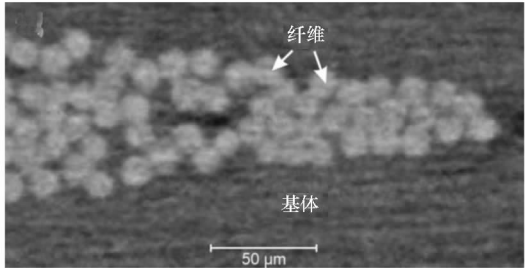
4.4 组分材料体分比的测定

4.4.1 全局纤维体分比 \widehat{V}_f 的测定

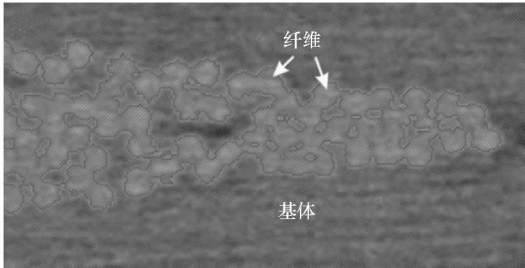
由于编织复合材料具有局部材质不均匀的特性,因此在测定全局纤维体分比时图像区域需要具有一定的尺度范围。一般来说,分析的显微 CT 图像区域越大,得到的全局纤维体分比 \widehat{V}_f 越准确。从图 9 可以看出,所获得的微尺度显微 CT 图像区域共涉及 3 个铺层,且每个铺层在长度方向上约四分之一周期长,其中涉及纤维束在相邻交叉点间的各个区段。这保证了全局纤维体分比 \widehat{V}_f 测定值的可靠性。

对图 9 所示区域的三维显微 CT 图像进行组分材料的分割并进行组分材料体分比的测定。由于聚合基复合材料的孔隙率低,故忽略图像区域中的孔隙,并将该图像区域分割为纤维和基体两种组分材料。由于显微 CT 图像中纤维和基体灰度值交叠较为严重,直接采用灰度阈值分割算法来对图像中的纤维和基体进行分割会在纤维边缘产生许多不规则的“毛刺”且单根纤维截面并不是圆形(如图 10(b)所示)(实际玻璃纤维是圆形截面,如图 2 所示)。这种误分割结果会严重影响全局纤维体分比的测定值。

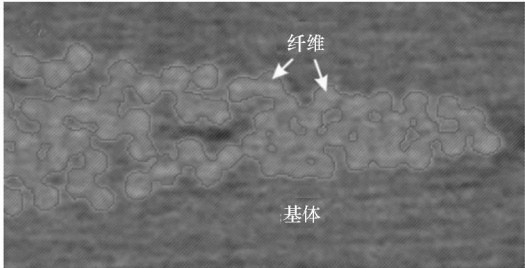
为了得到更加真实的区域分割结果,采用分水岭算法对图 9 所示的显微 CT 图像进行组分材料分割。图 10 是微尺度三维显微 CT 图像(图 9)中的某一纤维束局部图像的二维切片图。对图 10 (a)所示的纤维束局部区域进行分水岭算法分割结果如图 10 (c)所示。图 10 (c)的分割结果较图 10(b)有了明显改进,纤维边缘的“毛刺”现



(a) 纤维束局部显微 CT 图像
(a) Micro CT image of local yarn area



(b) 灰度阈值分割结果
(b) Result of gray level threshold method



(c) 分水岭算法结果
(c) Result of watershed segmentation method

图 10 纤维束局部显微 CT 图像的分割
Fig. 10 Segmentation of micro CT image of local yarn area

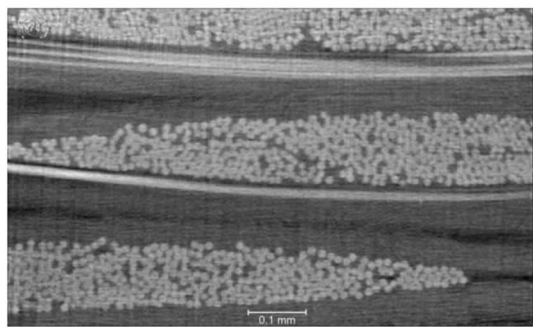
象有了明显抑制,且玻璃纤维截面近似圆形。

图 11(a)是对应图 9 中的一个 XZ 切片图,而图 11(b)是采用分水岭算法对图 9 进行组分材料分割后对应于图 11(a)的分割结果。通过分别测量图 9 中两种材料区域的体积可以获得全局纤维体分比 $\widehat{V}_f=0.522$ 。

4.4.2 局部纤维体分比 V_f 的测定

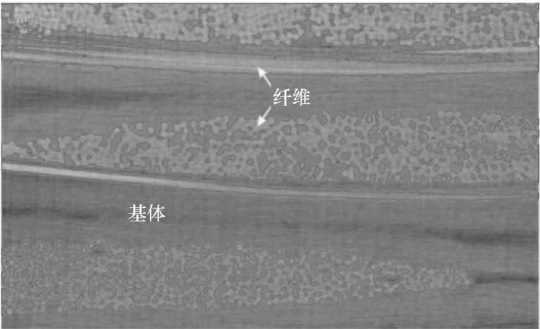
由于纤维束在织物中不同位置受到不同程度的挤压作用,其局部纤维体分比 V_f 随纤维束位置的变化而变化。因此,在测定局部纤维体分比 V_f 时,图像区域需要具有一定的尺度范围。受实验条件限制,选定与测定全局纤维体分比 \widehat{V}_f 一致的微尺度显微 CT 图像区域(图 9)进行局部纤维体分比 V_f 的测定。

在测定前,需要对该图像区域中的纤维束和基体区域进行剥离,以保证只对纤维束区域进行



(a) 微尺度显微 CT 图像的 XZ 切片

(a) XZ slice image of micro CT image in micro-scale



(b) 分水岭算法结果

(b) Result of watershed segmentation method

图 11 测定全局纤维体分比 \hat{V}_f 时组分材料分割结果的 XZ 切片图

Fig. 11 Resulting XZ slice image of constituent materials segmentation when measuring global fiber volume fraction \hat{V}_f

局部纤维体分比 V_f 测定。对图 9 所示的微尺度显微 CT 图像中纤维束间的基体富集区(即中尺度微观结构中的基体区域)进行剥离。由于纤维束间的基体富集区与纤维束内部的基体灰度值一致,因此在剥离基体富集区时,难以对二者进行灰度阈值分割。

为了解决该问题,采用图像形态学处理中的闭合操作对显微 CT 图像进行处理。闭合操作类似于先膨胀后侵蚀操作,但是该操作对目标对象近乎无损。该操作可以将小的目标对象包含入大的目标对象,同时填充内部的孔隙。此外,该操作还通过光滑边界和连通闭合的对象来清除小的细节。通过闭合操作生成对应的标签图像。将原始显微 CT 图像与标签图像作分割操作,可以将纤维束间的基体富集区在灰度值区间上分割开。剥离纤维束间的基体富集区后的微尺度显微 CT 图像如图 12 所示。从图 12 中可以看出,处于纤维束内部区域的基体被保留下来,而纤维束外部区域的基体被剥离。

图 13(a)是对应图 9 中位于 Y 轴中部的一个

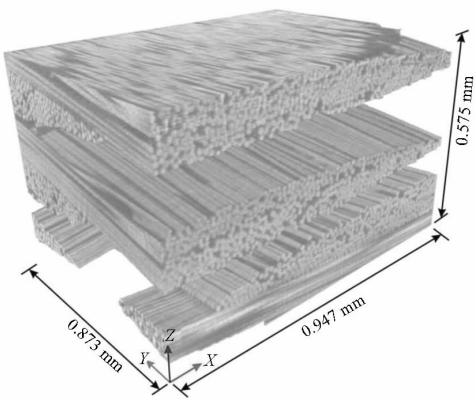
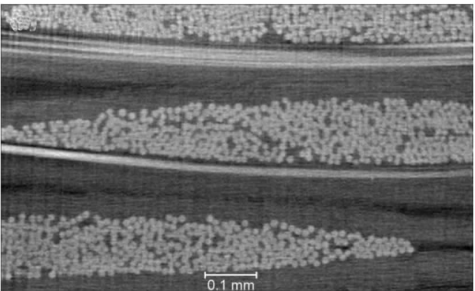


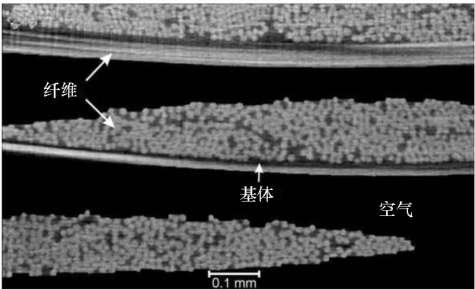
图 12 测定局部纤维体分比 V_f 的显微 CT 图像区域

Fig. 12 Micro CT image of measuring local fiber volume fraction V_f



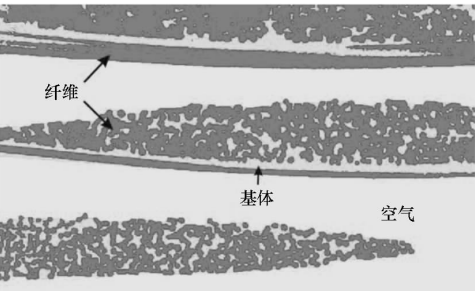
(a) 微尺度显微 CT 图像的 XZ 切片

(a) XZ slice image of micro CT image in micro-scale



(b) 剥离纤维束间基体富集区后的结果

(b) Result stripped matrix-enriched area between yarns



(c) 三组分材料分水岭算法结果

(c) Result of watershed segmentation method of three constituent materials

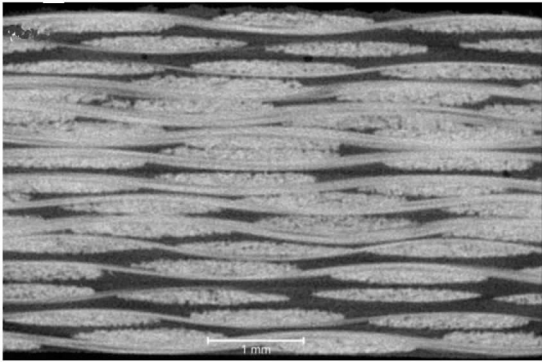
图 13 测定局部纤维体分比 V_f 时三组分材料分割结果的 XZ 切片图

Fig. 13 Resulting XZ slice image of three constituent materials segmentation when measuring local fiber volume fraction V_f

XZ 切片。图 13(b) 是图 13(a) 剥离纤维束间的基体富集区后的切片图像。在该图像中存在三个组分, 分别为纤维、基体和空气(背景)。对其进行分割时需要针对三组分材料进行分水岭分割。图 13(c) 是采用三组分材料分水岭算法对图 13(b) 进行分割后的结果。从图中可以看出, 经过一系列剥离和分割操作, 成功区分了纤维束中的纤维和基体。通过计算图 12 中纤维和基体区域的体积可以测得局部纤维体分比 $V_f=0.787$ 。

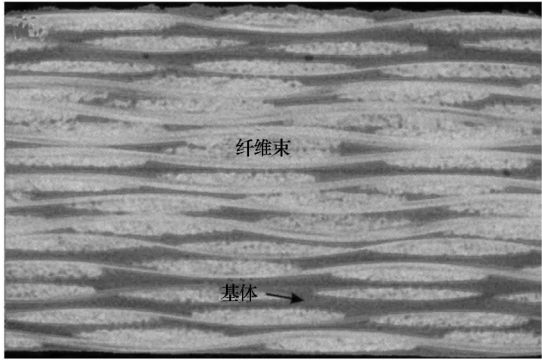
4.4.3 纤维束体分比 V_y 的测定

对图 8 所示的中尺度显微 CT 图像进行纤维束体分比 V_f 的测定。仍然采用分水岭算法对图像进行纤维束与基体分割, 分割结果如图 14 所示。通过测量纤维束在该区域中的体积占比可以获得纤维束体分比 $V_y=0.717$ 。



(a) 中尺度显微 CT 图像的 YZ 切片

(a) YZ slice image of micro CT image in meso-scale



(b) 分水岭算法结果

(b) Result of watershed segmentation method

图 14 测定纤维束体分比 V_y 时组分材料分割结果的 YZ 切片图

Fig. 14 Resulting YZ slice image of constituent materials segmentation when measuring yarn volume fraction V_y

5 讨论

表 3 罗列了三种实验测定法对 E-Glass/Epoxy 纤维织物复合材料组分材料体积分数的测

定值和相对误差。D3171 G 实验仅能测定全局纤维体分比 \tilde{V}_f , 扫描电镜实验仅能近似测定局部纤维体分比 V_f , 而不同尺度的显微 CT 实验可以同时测定全局纤维体分比 \tilde{V}_f 、局部纤维体分比 V_f 和纤维束体分比 V_y 。E-Glass/Epoxy 纤维织物复合材料实例表明, 利用不同尺度的显微 CT 图像来测定编织复合材料的组分材料体分比是可行的。此外, 显微 CT 实验测定法还可以为陶瓷基复合材料(如 C/SiC)等难以用常规物理实验(如化学消蚀法、D3171 G)测定体分比的复合材料组分材料体积分数测定提供解决方案。根据显微 CT 实验测定法的基本原理可以看出, 显微 CT 实验测定法还可以用于测定多孔复合材料的孔隙率。

从表 3 可以看出, 分别通过 D3171 G 实验和显微 CT 实验测定的全局纤维体分比 \tilde{V}_f 的相对误差为 2.353%, 这意味着通过显微 CT 图像测定 E-Glass/Epoxy 纤维织物复合材料的组分材料体积分数是合理的。显微 CT 实验测定法的误差主要来源于以下三个方面:

- 1) 所观测图像区域的尺度大小;
- 2) 图像分辨率;
- 3) 图像处理造成的误差。

这三个方面是造成与 D3171 G 实验测定值间误差的主要因素。

局部纤维体分比 V_f 的扫描电镜测定值与显微 CT 测定值的相对误差为 -4.828%, 这说明了对于 E-Glass/Epoxy 纤维织物复合材料而言, 局部纤维体分比 V_f 的扫描电镜测定值是可以接受的。

表 3 组分材料体积分数的三种实验测定结果

Tab. 3 Constituent contents measured by three experiment methods

	\tilde{V}_f	V_f	V_y
D3171 G 实验	0.510		
SEM 实验		0.749	
显微 CT 实验	0.522	0.787	0.717
相对误差/%	2.353 ^a	-4.828 ^b	

注:(·)^a=(显微 CT 实验值 - D3171 G 实验值) × 100/ D3171 G 实验值;

(·)^b=(SEM 实验值 - 显微 CT 实验值) × 100/ 显微 CT 实验值。

6 结论

1) D3171 G 实验只能获得纤维织物复合材料的全局纤维体分比 \tilde{V}_f ; 扫描电镜实验只能近似获

得纤维织物复合材料的局部纤维体分比 V_f ; 通过不同尺度的显微CT实验可以分别测得纤维织物复合材料的全局纤维体分比 \tilde{V}_f 、局部纤维体分比 V_f 以及纤维束体分比 V_y 。

2) 通过对比 E-Glass/Epoxy 纤维织物复合材料组分材料体分比的三种实验测定值, 说明了显微CT实验测定法的可行性和合理性。

3) 对于扫描电镜实验测定法和显微CT实验测定法分别给出了相应的图像处理方案。

4) 显微CT实验测定法可以为难以用常规物理实验测定体分比的复合材料组分材料体积分数测定提供解决方案。

参考文献 (References)

[1] Desplentere F, Lomov S, Woerdeman D, et al. Micro-CT characterization of variability in 3D textile architecture [J]. Composites Science and Technology, 2005, 65(13): 1920 – 1930.

[2] Madra A, Hajj N E, Benzeggagh M. X-ray microtomography applications for quantitative and qualitative analysis of porosity in woven glass fiber reinforced thermoplastic [J]. Composites Science and Technology, 2014, 95: 50 – 58.

[3] Pazmino J, Carvelli V, Lomov S. Micro-CT analysis of the internal deformed geometry of a non-crimp 3D orthogonal weave E-glass composite reinforcement [J]. Composites Part B; Engineering, 2014, 65(10): 147 – 157.

[4] Schell J S U, Renggli M, Van Lenthe G, et al. Micro-computed tomography determination of glass fibre reinforced polymer meso-structure [J]. Composites Science and Technology, 2006, 66(13): 2016 – 2022.

[5] Bale H, Blacklock M, Begley M R, et al. Characterizing three-dimensional textile ceramic composites using synchrotron X-ray micro-computed-tomography [J]. Journal of the American Ceramic Society, 2012, 95(1): 392 – 402.

[6] Vanaerschot A, Cox B N, Lomov S V, et al. Stochastic framework for quantifying the geometrical variability of laminated textile composites using micro-computed tomography [J]. Composites Part A: Applied Science and Manufacturing, 2013, 44: 122 – 131.

[7] Wang H, Wang Z W. Statistical analysis of yarn feature parameters in C/Epoxy plain-weave composite using micro CT with high-resolution lens-coupled detector [J]. Applied Composite Materials, 2016, 23(4): 1 – 22.

[8] Wang H, Wang Z W. A variable metric stochastic theory of textile composites with random geometric parameters of yarn cross-section [J]. Composite Structures, 2015, 126: 78 – 88.

[9] Wang H, Wang Z W. Quantification of effects of stochastic feature parameters of yarn on elastic properties of plain-weave composite—Part 1: Theoretical modeling [J]. Composites Part A: Applied Science and Manufacturing, 2015, 78: 84 – 94.

[10] Wang H, Wang Z W. Quantification of effects of stochastic feature parameters of yarn on elastic properties of plain-weave composite—Part 2: Statistical predictions vs. mechanical experiments [J]. Composites Part A: Applied Science and Manufacturing, 2016, 84: 147 – 157.

[11] Tan P, Tong L Y, Steven G P. Micromechanics models for the elastic constants and failure strengths of plain weave composites [J]. Composite Structures, 1999, 47(1/2/3/4): 797 – 804.

[12] Prodromou A G, Lomov S V, Verpoest I. The method of cells and the mechanical properties of textile composites [J]. Composite Structures, 2011, 93(4): 1290 – 1299.

[13] Cox B N, Flanagan G. Handbook of analytical methods for textile composites; NASA-CR-4750[R]. US: NASA Technical Reports Server, 1997.

[14] Lee S K, Byun J H, Hong S H. Effect of fiber geometry on the elastic constants of the plain woven fabric reinforced aluminum matrix composites [J]. Materials Science and Engineering: A, 2003, 347(1/2): 346 – 358.

[15] 朱元林, 崔海涛, 温卫东, 等. 含纤维束截面形状变化的三维编织复合材料细观模型及刚度预报 [J]. 复合材料学报, 2012, 29(6): 187 – 196.

ZHU Yuanlin, CUI Haitao, WEN Weidong, et al. Microstructure model and stiffness prediction of 3D braided composites considering yarns' cross-section variation [J]. Acta Materiae Compositae Sinica, 2012, 29(6): 187 – 196. (in Chinese)

[16] 张超, 许希武. 二维二轴编织复合材料几何模型及弹性性能预测 [J]. 复合材料学报, 2010, 27(5): 129 – 135.

ZHANG Chao, XU Xiwu. Geometrical model and elastic properties prediction of 2D biaxial braided composites [J]. Acta Materiae Compositae Sinica, 2010, 27(5): 129 – 135. (in Chinese)

[17] Standard test methods for constituent content of composite materials; ASTM D3171 – 99[S]. US: ASTM International, 2000.

[18] Lomov S V, Ivanov D S, Verpoest I, et al. Meso-FE modeling of textile composites: road map, data flow and algorithms [J]. Composites Science and Technology, 2007, 67(9): 1870 – 1891.

[19] 须颖, 邹晶, 姚淑艳. X 射线三维显微镜及其典型应用 [J]. CT 理论与应用研究, 2014, 24(6): 967 – 977.

XU Ying, ZOU Jing, YAO Shuyan. 3D X-ray microscope and its typical applications [J]. CT Theory and Applications, 2014, 24(6): 967 – 977. (in Chinese)

基于层裂机理的弹体侵彻混凝土的工程模型*

薛建锋^{1,2}, 沈培辉¹, 王晓鸣¹

(1. 南京理工大学 智能弹药技术国防重点学科实验室, 江苏 南京 210094;
2. 中航工业洪都 660 所, 江西 南昌 330024)

摘要:以斜侵彻过程中的终点弹道为研究对象,基于动态球形空腔膨胀理论给出的阻力函数理论公式和开坑阶段的表面层裂机理,建立了能够综合考虑弹头形状、开坑区深度的斜侵彻深度预测模型,并进一步推导了能够适用不同弹头形状的弹体过载时程曲线计算公式。预测模型得到的侵彻深度和过载与试验结果吻合较好。研究结果可为弹体与混凝土靶的斜侵彻弹道分析和弹丸头部设计提供一定帮助。

关键词:爆炸力学;层裂机理;终点弹道;侵彻深度;过载

中图分类号: O385; TJ012.4 **文献标志码:** A **文章编号:** 1001-2486(2017)03-194-07

Engineering model of projectile penetrating into concrete based on splitting mechanism

XUE Jianfeng^{1,2}, SHEN Peihui¹, WANG Xiaoming¹

(1. ZNDY Ministerial Key Laboratory, Nanjing University of Science and Technology, Nanjing 210094, China;
2. 660 Design Institute of Hong Du AVIC, Nanchang 330024, China)

Abstract: Based on the theoretical resistance function from dynamic spherical cavity expansion model and the surface splitting mechanism, a prediction model was proposed for oblique penetration depth of rigid projectiles into concrete targets. In the proposed formula, the dimensionless coefficients denoting the projectile nose geometry were introduced to consider the variation of projectile nose geometry and the cratering depth. A prediction model for the deceleration-time history of the projectile with different nose geometries was obtained. The penetration depths and overload curve of the model were in good agreement with the test results. The research results can provide some help for the oblique penetration trajectory analysis and the design of the projectile head.

Key words: explosion mechanics; splitting mechanism; terminal ballistic; penetration depth; overload

弹体侵彻混凝土靶的终点弹道研究一直是武器研发部门和防护工程人员的研究重点^[1]。弹体冲击混凝土靶的侵彻深度及过载时程是主要研究内容,其中侵彻深度是表征靶体破坏效应的最主要参数,过载时程的预测对弹体的壁厚及装药安定性设计具有指导意义。弹体斜侵彻混凝土介质时,由于靶体对弹体的阻力的不对称影响导致偏转,弹体的侵彻能力下降,侵彻深度减小,侵彻靶体的弹道会有一定量的偏转,并在一定条件下出现弹体在靶板表面跳飞现象^[2-6]。

基于阻力函数建立的半理论半经验公式能同时预测弹体侵彻深度及过载时程。Forrestal等^[7-9]基于空腔膨胀理论和弹体侵彻深度试验数据求得了半理论半经验的阻力函数,建立了

尖卵形弹体的侵彻深度和过载时程曲线预测模型;Chen Li等^[10-12]引入弹形系数和撞击函数,将Forrestal侵彻深度公式无量纲化,使得公式能够适用几种典型弹头形状弹体;Wen等^[13-14]根据由准静态压力和动压力两部分组成的阻力函数,建立了考虑弹头形状的弹体侵彻深度预测公式;Jan等^[15]、吴昊等^[16]和彭永等^[17]分别基于不同空腔膨胀理论的阻力函数建立了侵彻深度预测模型。然而,已有的研究多基于半经验半理论阻力函数,且没有考虑开坑阶段的表面层裂效应。对于过载时程预测模型,仅Forrestal基于半经验半理论阻力函数对卵形弹体的过载时程有所研究,针对其他弹体形状的研究未见报道。

* 收稿日期:2016-01-27
基金项目:国家重点基础研究发展计划资助项目(613143020301)
作者简介:薛建锋(1987—),男,江西抚州人,博士研究生,E-mail:xuejianfeng666@163.com;
沈培辉(通信作者),男,教授,硕士,硕士生导师,E-mail:sphjy8@mail.njust.edu.cn

1 斜侵彻深度的预测模型

1.1 弹体头部形状描述

对卵形和锥形弹体进行分析,如图 1 所示。

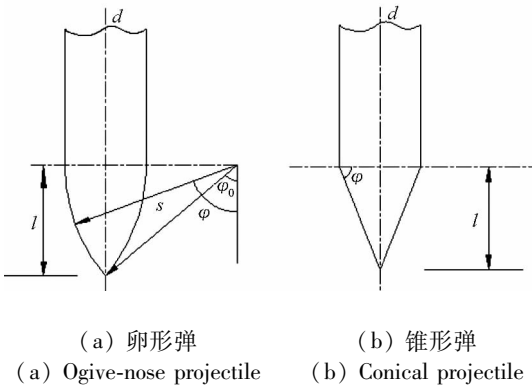


图 1 不同弹体头部几何示意图

Fig. 1 Geometry of projectiles with different nose

对于卵形头部弹体,直径为 d ,头部长度为 l ,头部曲率半径为 s ,则对应的描述弹形的各参数^[10]为:

$$\psi = \frac{s}{d} \tag{1}$$

$$\zeta = \frac{l}{d} \tag{2}$$

$$\phi = \arcsin\left(1 - \frac{1}{2\psi}\right), \quad \psi \geq \frac{1}{2} \tag{3}$$

$$\psi^2 = \zeta^2 + \left(\psi - \frac{1}{2}\right)^2 \tag{4}$$

$$N_0 = \zeta^2 - \frac{2\psi - 1}{2} \tag{5}$$

$$N_1 = \psi^2 \left\{ \frac{2}{3} \cdot \frac{\zeta^3}{\psi^3} + \frac{2\psi - 1}{2\psi} \left[\frac{(2\psi - 1)\zeta}{2\psi^2} + \arccos\left(\frac{\zeta}{\psi}\right) - \frac{\pi}{2} \right] \right\} \tag{6}$$

$$N_2 = 2\psi^2 \left\{ \frac{1}{4} \cdot \frac{\zeta^4}{\psi^4} + \frac{2\psi - 1}{2\psi} \left[\frac{(2\psi - 1)}{2\psi} + \frac{(2\psi - 1)^3}{24\psi^3} - \frac{2}{3} \right] \right\} \tag{7}$$

式中: N_0 为表征弹体头部形状的形状因子; ψ 表示卵形弹体头部的曲径比; N_1 和 N_2 为与头部形状有关的无量纲形状系数。

对于锥形头部弹体,直径为 d ,则对应的描述弹形的各参数^[10]为:

$$N_0 = \frac{1}{4} \tag{8}$$

$$N_1 = \frac{1}{4 \times (4\zeta^2 + 1)^{1/2}} \tag{9}$$

$$N_2 = \frac{1}{4 \times (4\zeta^2 + 1)} \tag{10}$$

定义撞击函数 I 和弹体几何函数 N ,具体表达式^[10]为:

$$I = \frac{MV_0^2}{d^3 YAN_1} \tag{11}$$

$$N = \frac{M}{\rho d^3 BN_2} \tag{12}$$

式中: ρ 和 Y 分别为混凝土的密度和屈服强度; M 和 V_0 分别为弹体的质量和速度; A 和 B 为待定系数,可通过实验确定。由上可知,弹体头部越尖细, I 和 N 越大,这将导致终点弹道特性产生差异。以上关于侵彻深度和偏转角的计算仅限于刚性弹体。

1.2 层裂机理

弹体斜侵彻混凝土过程中产生的应力波斜入射自由表面,入射波在自由表面反射出膨胀波和剪切波,并且这两种反射波的强度分配与入射角有关,剪切波对侵彻过程的影响程度小,在此忽略不计。

某一时刻平面压缩波倾斜入射到自由表面,在自由表面某一位置入射波波头阵面先到达,反射成为反射平面波波头阵面;入射压缩波后续阵面陆续到达,反射成为反射平面波后续阵面。随着时间的进展,在自由表面这个位置将进行这个平面波的入射和反射过程。图 2 是某时刻在自由表面处平面波入射和反射的图像, MN 为自由表面。

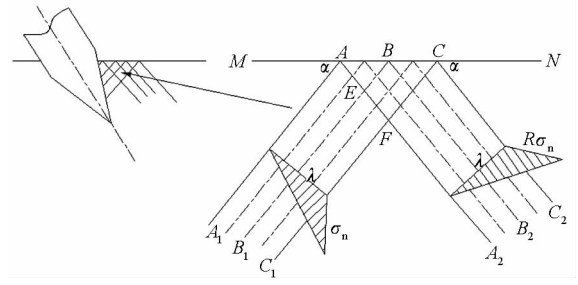


图 2 某时刻平面波在自由表面的入射和反射图像

Fig. 2 Geometry of incident and reflection of plane wave at free surface

设入射平面波波长为 λ ,强度为 σ_n ,入射角为 α 。按照反射波和入射波的关系可知:反射波的波长为 λ ,反射角为 α ,强度为 $R\sigma_n$,其中 R 是反射系数,它是入射角 α 和材料泊松比 ν 的函数,可表示^[18]为:

$$R = \frac{\left[\frac{2(1-\nu)}{1-2\nu} - \sin^2 \alpha \right]^{1/2} \sin^3 \alpha - \left(\frac{1-\nu}{1-2\nu} - \sin^2 \alpha \right) \tan \alpha}{\left[\frac{2(1-\nu)}{1-2\nu} - \sin^2 \alpha \right]^{1/2} \sin^3 \alpha + \left(\frac{1-\nu}{1-2\nu} - \sin^2 \alpha \right) \tan \alpha} \tag{13}$$

入射平面波和反射平面波相交的区域内,各

点的应力状态为该点处入射波强度和反射波强度按具体相交角度进行叠加。图 2 中 E 点处应力为反射波波头强度和入射波 $1/2$ 波长处的强度按相交角度 $180^\circ - 2\alpha$ 进行叠加; F 点处应力为反射波波头强度。任选一点 G (如图 3 所示), 设 G 距入射平面波波头阵面的距离为 ξ ; G 点处入射平面波强度为 $\sigma_I = (1 - \xi/\lambda)\sigma_n$, 其方向沿入射平面波波阵面法线方向; G 点处反射平面波强度为 $\sigma_{II} = R\sigma_n$, 其方向沿反射平面波波阵面法线方向; 二应力之间的夹角为 $180^\circ - 2\alpha$ 。按照纵波叠加主应力公式, 可以计算 σ_I, σ_{II} 叠加后的主应力为:

$$\sigma_c = \frac{\sigma_n}{2(1-\nu)} \left\{ 1 - \frac{\xi}{\lambda} + R + (1-2\nu) \cdot \left[\left(1 - \frac{\xi}{\lambda} \right)^2 + R^2 \left(1 - \frac{\xi}{\lambda} \right) \cos(4\alpha) \right]^{1/2} \right\} \quad (14)$$

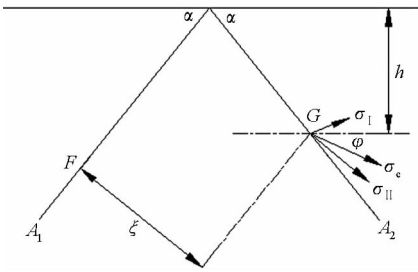


图 3 反射波与入射波的叠加

Fig. 3 Superposition of reflection wave and incident wave

叠加后主应力方向确定, 二应力方向夹角平分线恰好与自由表面平行, 主应力与该平行线的夹角, 即主应力与自由表面的夹角, 设该角为 φ , 表示为:

$$\tan(2\varphi) = \frac{\sigma_{II} - \sigma_I}{\sigma_{II} + \sigma_I} \tan(2\alpha) \quad (15)$$

将 σ_I 和 σ_{II} 的计算公式代入式(15)可得:

$$\tan(2\varphi) = \frac{1 - \xi/\lambda - R}{1 - \xi/\lambda + R} \tan(2\alpha) \quad (16)$$

根据式(14)求得的主应力为拉应力, 且最大拉应力大于混凝土的抗拉强度极限, 这时在混凝土表面将出现层裂, 层裂方向垂直于最大拉应力方向。层裂位置与自由表面的距离为:

$$h = \frac{\xi}{2\cos\alpha} \quad (17)$$

1.3 侵彻深度模型

假设弹体的侵彻过程始终在射平面内, 弹体对混凝土靶的斜侵彻过程分为开坑段和隧道段两部分, 弹体偏转发生在开坑阶段, 如图 4 所示。着靶时, 弹体侵彻弹道与靶面法线夹角为 β 。开坑段结束时, 弹体发生了偏转, 偏转角为 δ 。

文献[8, 19]表明: 正侵彻时, 开坑区深度与

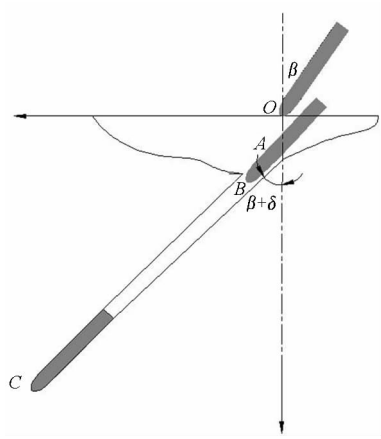


图 4 斜侵彻过程

Fig. 4 Oblique penetration

弹体直径成正比, 为 kd , k 为比例系数。Forrestal 等^[9]基于一系列试验数据得出比例系数为 2; Li 等^[12]基于滑移线理论, 得出可适应于不同弹头形状的比例系数为 $0.707 + l/d$ 。斜侵彻时, 比例系数为 $(0.707 + l/d) \cos\beta$ 。

在开坑区, 即 $x < kd$ 时, 弹体阻力表达式^[8]为:

$$F = cx \quad (18)$$

式中: c 为待定阻力系数, 与速度密切相关, 需根据连续条件求解。

根据牛顿第二定律及式(18)可知, 弹体在开坑区的运动方程为:

$$F = M \frac{d^2x}{dt^2} = cx \quad (19)$$

式(19)也可表示为:

$$F = M \frac{dV}{dx} \cdot \frac{dx}{dt} = MV \frac{dV}{dx} = cx \quad (20)$$

根据式(20)可得开坑阶段时侵彻深度为:

$$x = \left(\frac{MV_0^2}{2c} \right)^{1/2} \quad (21)$$

式中: V_0 为弹体着靶速度, c 取值^[8]一般为

$$c = \frac{\pi d^2 S f_c}{4k} \left[\frac{1 + I/N}{1 + (k\pi/4N)} \right] \quad (22)$$

其中: S 为混凝土静态阻力系数, f_c 为混凝土材料的单向无侧限抗强度。

由 $x(t=0) = 0$ 和 $x'(t=0) = V_0$, 解式(19)微分方程可得弹体在开坑区的位移时程表达式(23), 依次求导可得弹体速度、加速度方程。

$$x(t) = \frac{1}{\omega} V_0 \sin(\omega t) \quad (23)$$

$$V(t) = V_0 \cos(\omega t) \quad (24)$$

$$a(t) = -\omega V_0 \sin(\omega t) \quad (25)$$

其中: $\omega = \sqrt{c/M_0}$ 。

弹体偏转角 δ 可由动能定理求得。弹体垂直于弹轴的速度为:

$$V_{\perp} = V_0 \sin \delta \tag{26}$$

此时偏转的距离为:

$$s_{\perp} = x \delta \tag{27}$$

试验表明阻力随时间近似呈线性增加,因此假设平均侧向力为:

$$F_{\perp \text{ avg}} = \frac{1}{2} F_0 \sin \beta = \frac{1}{2} c x \sin \beta \tag{28}$$

式中: F_0 可由式(18)确定。根据动能定理得:

$$\frac{1}{2} M V_{\perp}^2 = F_{\perp \text{ avg}} s_{\perp} \tag{29}$$

最终求得偏转角为:

$$\sin^2 \delta = \delta \frac{k \pi}{4} \left(\frac{1}{I} + \frac{1}{N} \right) \sin \beta \tag{30}$$

在隧道区,根据动态球形空腔膨胀理论^[7]和层裂机理,可将空腔膨胀应力与空腔膨胀速度的关系表达为:

$$\sigma_r = R [A f_c + B (\rho f_c)^{1/2} V_r + C \rho V_r^2] \tag{31}$$

式中: A, B 和 C 为材料系数;空腔膨胀速度与弹体速度满足 $V_r = V \cos \theta$; R 由式(13)可得。

将式(31)沿不同形状弹体头部表面积分可得到阻力函数。基于文献[20]的研究工作,在终点弹道行为研究中,弹体摩擦力忽略不计。弹体轴向曲面受力为:

$$F = - \iint_{\Sigma} \sigma_r \cos \theta d\sigma \tag{32}$$

式中: Σ 为曲面的表面积。

由式(31)和式(32)可得到弹体在隧道区的阻力函数为:

$$F = - \pi d^2 R [N_0 A f_c + N_1 B (\rho f_c)^{1/2} V + N_2 C \rho V^2] \tag{33}$$

式中: N_0, N_1 和 N_2 为弹形系数,按 1.1 节公式计算。

定义开坑区结束时的速度 V_1 和时刻 t_1 ,将式(20)积分可得:

$$M (V_1 - V_0) / 2 = c (k d)^2 / 2 \tag{34}$$

根据开坑区与隧洞段交界面弹体所受阻力连续,可得:

$$c k d = R \pi d^2 [N_0 A f_c + N_1 B (\rho f_c)^{1/2} V_1 + N_2 C \rho V_1^2] \tag{35}$$

联立式(18)和式(19)可得到 V_1 ,同时根据式(24)可以求得 t_1 。

$$t_1 = \frac{\arccos(V_1/V_0)}{\omega} \tag{36}$$

隧道区的运动方程可根据牛顿第二定律和

式(33)求得。

$$F = M \frac{dV}{dt} = M V \frac{dV}{dx} = - \pi d^2 R [N_0 A f_c + N_1 B (\rho f_c)^{1/2} V + N_2 C \rho V^2] \tag{37}$$

将式(37)积分联立初始条件就可以求得侵彻深度为:

$$x = \frac{R M}{2 \pi C N_2 \rho d^2} \left\{ \ln \left[1 + \frac{B N_1}{A N_0} \left(\frac{\rho}{f_c} \right)^{1/2} V_1 + \frac{C N_2 \rho}{A N_0 f_c} V_1^2 \right] + \frac{2 B N_1}{D} \left[\arctan \left(\frac{B N_1}{D} \right) - \arctan \left(\frac{B N_1 + 2 C (\rho / f_c)^{1/2} N_2 V_1}{D} \right) \right] \right\} + k d \tag{38}$$

式中: $D = (4 A C N_0 N_1 - B^2 N_1^2)^{1/2}$ 。

通过式(37)也可以求得隧道区的速度时程表达式为:

$$V(t) = \frac{R}{2 C N_2 (\rho / f_c)^{1/2}} \cdot \left\{ D \tan \left[\arctan \left(\frac{2 C N_2 (\rho / f_c)^{1/2} V_1 + B N_1}{D} \right) - \frac{\pi d^2}{2 M} D (\rho f_c)^{1/2} (t - t_1) \right] - B N_1 \right\} \tag{39}$$

对速度时程表达式(39)一次求导,可得弹体在隧道区的加速度时程为:

$$a(t) = \frac{- R \pi d^2 D f_c}{4 M C N_2 \cos^2 \mathcal{K}} \tag{40}$$

$$\mathcal{K} = \arctan \left[\frac{2 V_1 C N_2 (\rho / f_c)^{1/2} + B N_1}{D} \right] \tag{41}$$

2 试验验证

2.1 试验弹体和混凝土靶

试验弹体为两种结构类型,即卵形和锥形,其中卵形弹的头部形状系数(Caliber Radius Head, CRH)分别为 3 和 4,在保证弹体质量不变的情况下,弹体直径为 10 mm,长径比为 7,质量约为 105 g。锥形弹直径为 10 mm,锥角为 15°。通过改变弹体内部孔洞的长度来保证弹体质量相等。所有弹体材料为高强度钢 35CrMnSiA,淬火处理后其屈服强度为 1500 MPa,硬度值为 45。为了消除靶体侧面边界效应的影响,靶体直径取 500 mm,即弹体直径的 50 倍。为了近似半无限厚靶处理,靶体厚度需要够大,厚度取 300 mm。为了方便浇注混凝土靶并保持侵彻后靶体的完整性,靶板外围采用 3 mm 厚的钢圈加固,混凝土靶面浇注成斜置 30°。浇注三个长宽高均为 100 mm 的混凝土样品,抗压测试强度为 36.2 MPa。

2.2 侵彻试验方法与结果

以 25 mm 口径的滑膛炮作为发射平台进行弹体侵彻混凝土靶试验,试验现场布置如图 5 所示,通过装药量控制弹体着靶速度,用锡箔靶和双通道测试仪测量速度。利用研制的加速度测试系统对过载进行测试,加速度测试系统包括加速度传感器、存储记录部分、数据接口、数据处理软件。将加速度过载测试仪安放在弹体内部孔洞里,并加配合垫片防止脱落。详细试验结果见表 1。

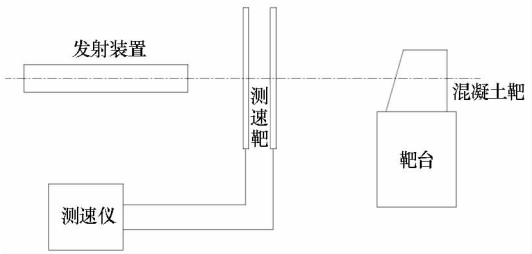


图 5 试验布置图
Fig. 5 Experiment layout

表 1 试验结果
Tab. 1 Experimental results

序号	头部形状	速度/(m/s)	开坑直径/mm	开坑深度/mm	侵彻深度/mm
1	卵形(CRH=4)	712	112	30.2	114
2	卵形(CRH=4)	805	125	35.6	143
3	卵形(CRH=4)	920	142	40.4	186
4	卵形(CRH=4)	1051	159	45.4	225
5	卵形(CRH=4)	1159	168	47.3	241
6	卵形(CRH=3)	704	107	29.8	109
7	卵形(CRH=3)	803	113	32.6	132
8	卵形(CRH=3)	924	134	37.2	176
9	卵形(CRH=3)	1048	146	41.7	204
10	卵形(CRH=3)	1152	158	43.5	235
11	锥形	708	102	28.5	102
12	锥形	806	102	30.1	128
13	锥形	923	119	35.2	172
14	锥形	1053	132	37.9	198
15	锥形	1161	141	40.7	229

图 6 为速度 805 m/s 下卵形弹侵彻后靶体正面破坏情况,崩落区的形状不是很规则,但基本上

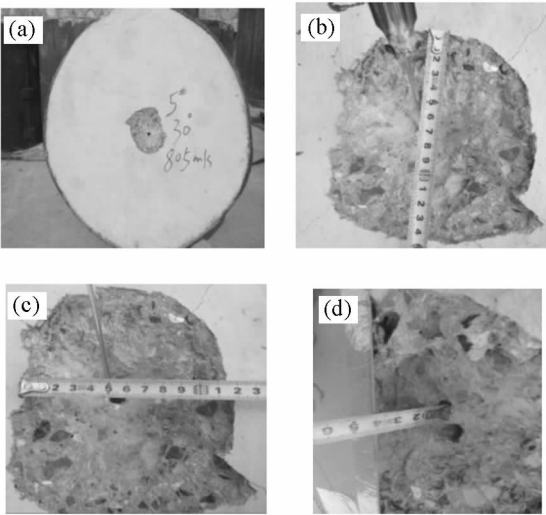


图 6 靶体破坏情况
Fig. 6 Damage effects of concretes

呈漏斗形。图 7 为速度 805 m/s 下卵形弹侵彻混凝土靶的隧道剖面,弹洞的直径经过测量稍微大于弹径,形状并不是理论意义上的圆柱形,而是稍微有点弯曲的,这是由于弹体在斜侵彻过程中存在一定的偏转。图 8 为回收的部分弹体,由于弹体后部分有螺纹或内部空心,在侵彻过程中弹体发生断裂或破裂。

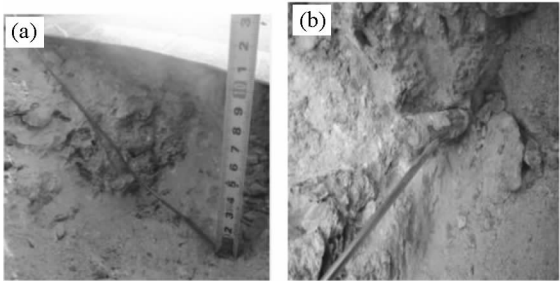


图 7 靶体剖开后隧道形态
Fig. 7 Tunnel morphology of cleaved target



图8 回收弹体
Fig.8 Recovered projectiles

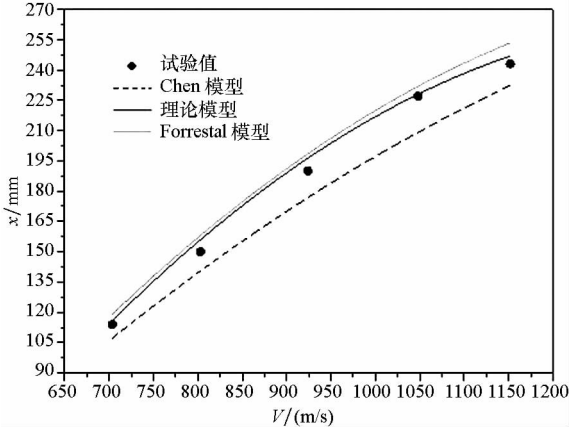
2.3 理论计算与试验结果对比分析

表2 为试验中卵形弹($CRH=4$)偏转角与理论计算值及Chen模型值的比较。从表2中可以看出,考虑层裂机理的模型得到的偏转角与试验结果较为接近。这是因为在开坑过程中一旦出现层裂,就形成了新的自由表面。继续入射的压缩波将在此新自由表面上反射,继续发生层裂。随着层裂的发生,弹体上下表面所受阻力的不对称性导致弹体发生偏转。

表2 偏转角的比较
Tab.2 Comparison of deflection angle

速度/(m/s)	本模型/(°)	试验值/(°)	Chen模型/(°)
1051	8.5	9.6	12.5
920	12.6	13.4	15.7
805	16.3	17.5	20.6

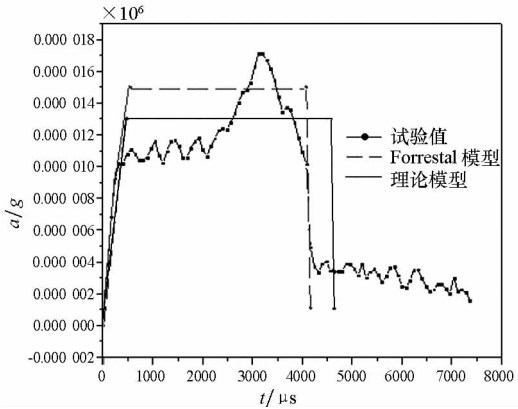
图9给出了本文侵彻深度计算公式、试验结果及几种半经验方法计算出的不同头部形状弹体斜侵彻混凝土靶体侵彻深度预测值的对比。计算中的参数值与试验中弹靶参数值一致,其中,混凝土靶密度为 2370 kg/m^3 ,强度为 36.2 MPa , A,B 和 C 分别为 $7.14,5.03,0.47$ 。可见提出的考虑弹形系数影响的侵彻深度计算模型预测值与试验



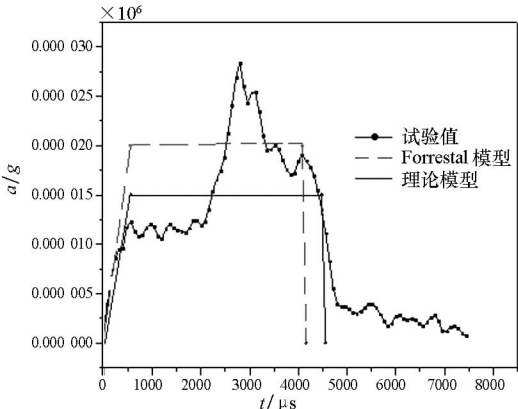
(b) 锥形弹
(b) Conical projectile

图9 侵彻深度试验值和公式预测结果对比
Fig.9 Comparison of projectile penetration depth of experiment and formula

侵彻深度吻合较好,证明了该计算模型的正确性。图10给出了加速度试验值、本文过载时程预

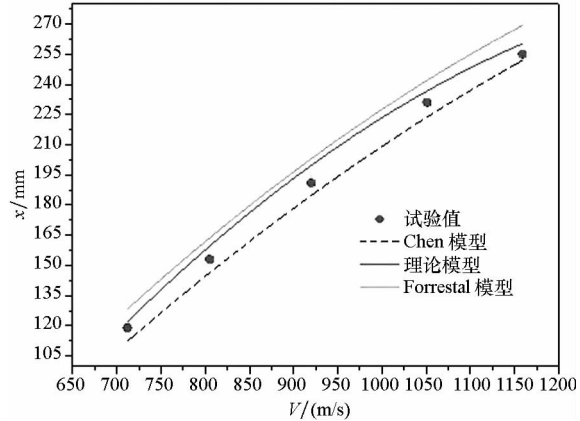


(a) 卵形弹(803 m/s)
(a) Ogive-nose projectile (803 m/s)



(b) 锥形弹(923 m/s)
(b) Conical projectile (923 m/s)

图10 加速度时程的试验值和公式预测结果对比
Fig.10 Comparison of acceleration time history of experimental values and formulas



(a) 卵形弹($CRH=4$)
(a) Ogive-nose projectile ($CRH=4$)

测模型和 Forrestal 计算模型^[9]的计算值。从图 10 中可以看出,本文模型计算出的加速度时程曲线与试验的吻合程度更好,在开坑初期加速度迅速增加后趋于平稳;当开坑结束弹道发生偏转,产生翻转力矩,此时加速度急速增加至极值,再大幅度减小至零。该加速度模型能较好地关注侵彻终点时刻(可用于引信的延迟时刻设计),并能更好地关注弹体加速度值(可用于弹体强度设计)。

3 结论

- 1) 通过与不同头形弹体侵彻深度试验数据以及过载时程曲线的对比,验证了计算公式的正确性。
- 2) 建立了适用于不同头部形状弹体的过载时程计算公式,为弹体稳定性、装药安定性及引信设计提供了理论基础。基于动态球形空腔膨胀理论得到的阻力函数理论公式在预测过载时程曲线方面比半经验半理论公式更加准确。

参考文献 (References)

[1] 张若棋, 汤文辉, 赵国民, 等. 长杆射弹侵彻三种混凝土靶的实验研究[J]. 国防科技大学学报, 2004, 26(5): 22-25.
ZHANG Ruoqi, TANG Wenhui, ZHAO Guomin, et al. Experimental study of the long rod projectile penetrating three kinds of concrete targets [J]. Journal of National University of Defense Technology, 2004, 26(5): 22-25. (in Chinese)

[2] 李进忠, 蔡汉文, 崔秉贵, 等. 混凝土侵彻的工程计算模型[J]. 兵工学报, 1995, 16(4): 86-88.
LI Jinzhong, CAI Hanwen, CUI Binggui, et al. An engineering calculation model for the penetration in concrete targets [J]. Acta Armamentarii, 1995, 16(4): 86-88. (in Chinese)

[3] 武海军, 黄风雷, 王一楠. 高速弹体非正侵彻混凝土试验研究[C]//第八届全国爆炸力学学术会议文集, 吉安, 2007: 488-494.
WU Haijun, HUANG Fenglei, WANG Yinan. Experimental research on high-velocity penetration into concrete targets[C]//Proceeding of the 8th Chinese Conference on Explosion Mechanics, Ji'an, 2007: 488-494. (in Chinese)

[4] 马爱娥, 黄风雷. 弹体斜侵彻钢筋混凝土的试验研究[J]. 北京理工大学学报, 2007, 27(6): 482-486.
MA Aie, HUANG Fenglei. Experimental research on oblique penetration into reinforced concrete [J]. Transactions of Beijing Institute of Technology, 2007, 27(6): 482-486. (in Chinese)

[5] 吕中杰, 徐钰巍, 黄风雷. 弹体斜侵彻混凝土过程中的方向偏转[J]. 兵工学报, 2009, 30(s2): 301-304.
LYU Zhongjie, XU Yuwei, HUANG Fenglei. Transverse deflection of projectile obliquely penetrating into concrete[J]. Acta Armamentarii, 2009, 30(s2): 301-304. (in Chinese)

[6] 王可慧, 宁建国, 李志康, 等. 高速弹体非正侵彻混凝土靶的弹道偏转试验研究[J]. 高压物理学报, 2013, 27(4): 561-566.
WANG Kehui, NING Jianguo, LI Zhikang, et al. Ballistic trajectory of high-velocity projectile obliquely penetrating

concrete target [J]. Chinese Journal of High Pressure Physics, 2013, 27(4): 561-566. (in Chinese)

[7] Forrestal M J, Luk V K. Dynamic spherical cavity-expansion in a compressible elastic-plastic solid [J]. Journal of Applied Mechanics, 1988, 55(2): 275-279.

[8] Forrestal M J, Tzou D Y. A spherical cavity-expansion penetration model for concrete targets [J]. International Journal of Solids and Structures, 1997, 34(31/32): 4127-4146.

[9] Forrestal M J, Altman B S, Gargile J D, et al. An empirical equation for penetration depth of ogive-nose projectiles into concrete targets [J]. International Journal of Impact Engineering, 1994, 15(4): 395-405.

[10] Chen X W, Li Q M. Deep penetration of a non-deformable projectile with different geometrical characteristics [J]. International Journal of Impact Engineering, 2002, 27(6): 619-637.

[11] Li Q M, Reid S R, Wen H M, et al. Local impact effects of hard missiles on concrete targets[J]. International Journal of Impact Engineering, 2005, 32(1/2/3/4): 224-284.

[12] Li Q M, Chen X W. Dimensionless formulae for penetration depth of concrete targets impacted by rigid projectiles [J]. International Journal of Impact Engineering, 2003, 28(1): 93-116.

[13] Wen H M. Predicting the penetration and perforation of targets struck by projectiles at normal incidence [J]. Mechanics of Structures and Machines, 2002, 30(4): 543-577.

[14] Wen H M, Yang Y. A note on the deep penetration of projectile into concrete [J]. International Journal of Impact Engineering, 2014, 66(4): 1-4.

[15] Teland J A, Sjol H. Penetration into concrete by truncated projectiles [J]. International Journal of Impact Engineering, 2004, 30(4): 447-464.

[16] 吴昊, 方秦, 龚自明, 等. 考虑刚性弹头形状的混凝土(岩石)靶体侵彻深度半理论分析[J]. 爆炸与冲击, 2012, 32(6): 573-580.
WU Hao, FANG Qin, GONG Ziming, et al. Semi-theoretical analyses for penetration depth of rigid projectiles with different nose geometries into concrete (rock) targets [J]. Explosion and Shock Waves, 2012, 32(6): 573-580. (in Chinese)

[17] 彭永, 方秦, 吴昊, 等. 不同头部形状弹体侵彻混凝土靶体的终点弹道参数分析[J]. 兵工学报, 2014, 35(s2): 128-134.
PENG Yong, FANG Qin, WU Hao, et al. Theoretical analyses for terminal ballistic of the projectiles with different nose geometries penetrating into concrete targets [J]. Acta Armamentarii, 2014, 35(s2): 128-134. (in Chinese)

[18] 宁建国. 爆炸与冲击力学[M]. 北京: 国防工业出版社, 2010: 298-300.
NING Jianguo. Explosion and impact mechanics [M]. Beijing: National Defense Industry Press, 2010: 298-300. (in Chinese)

[19] 石志勇, 汤文辉, 赵国民, 等. 长杆射弹对钢纤维混凝土靶开坑特性的实验研究[J]. 国防科技大学学报, 2004, 26(5): 26-29.
SHI Zhiyong, TANG Wenhui, ZHAO Guomin, et al. Experimental study of the crater performance about the long rod projectile penetrating steel fiber reinforced concrete target [J]. Journal of National University of Defense Technology, 2004, 26(5): 26-29. (in Chinese)

[20] Chen X W, Fan S C, Li Q M. Oblique and normal perforation of concrete targets by a rigid projectile [J]. International Journal of Impact Engineering, 2004, 30(6): 617-637.

COSMIC 掩星反演的水汽廓线质量分析*

陈志平¹, 罗佳^{1,2}, 肖晓¹, 孙方方¹

(1. 武汉大学测绘学院, 湖北武汉 430079;
2. 武汉大学地球空间环境与大地测量教育部重点实验室, 湖北武汉 430079)

摘要:利用 2006—2012 年北半球冬季低纬度地区(30°S~30°N)无线电探空站数据及全球大气成分和气候监测再分析数据对 1000 hPa~200 hPa 高度层的气象、电离层与气候星座观测系统全球定位系统掩星反演的比湿廓线进行了精度和可靠性验证。结果表明,水汽对气象、电离层与气候星座观测系统掩星反演影响较大,尤其在中、低对流层及热带地区等水汽含量比较大的地区,且气象、电离层与气候星座观测系统掩星数据在 850 hpa 以下可能并不太适用于评估其他数据。

关键词:气象、电离层与气候星座观测系统;GPS 掩星;比湿
中图分类号:P413 **文献标志码:**A **文章编号:**1001-2486(2017)03-201-06

Assessment of COSMIC radio occultation water vapor profile

CHEN Zhiping¹, LUO Jia^{1,2}, XIAO Xiao¹, SUN Fangfang¹

(1. School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, China;

2. Key Laboratory of Geospace Environment and Geodesy, Ministry of Education, Wuhan University, Wuhan 430079, China)

Abstract: The specific humidity profiles derived from the GPS (global positioning system) RO (radio occultation) of the COSMIC (constellation observing system for meteorology, ionosphere, and climate) with those from low latitude (30°S~30°N) radiosonde and MACC (monitoring atmospheric composition and climate) reanalysis during the NH (north hemisphere) winters (December, January, and February) from 2006 to 2012 over the layers from 1000 hPa to 200 hPa were verified. Comparison results demonstrate that the impact of water vapor on the COSMIC GPS RO inversion is very large, especially over the tropical and low troposphere. And the COSMIC GPS RO observations below 850 hpa may not be suitable for the assessment of other observations.

Key words: constellation observing system for meteorology, ionosphere, and climate; GPS radio occultation; specific humidity

美国和中国台湾地区合作研究的气象、电离层与气候星座观测系统 (Constellation Observing System for Meteorology, Ionosphere, and Climate, COSMIC) 计划是由 6 颗低轨卫星组成,于 2006 年 4 月 15 日 01:40 UTC 成功发射。COSMIC 是一种全球定位系统 (Global Positioning System, GPS) 掩星观测系统,具有全球覆盖、全天候、自校准且长期稳定、高垂直分辨率、高精度等优点^[1-3],对气象预报、气候监测和电离层等领域的研究具有广泛的应用^[4]。每天约有 2000 个掩星廓线均匀分布在大气中。COSMIC 使用了开环跟踪技术^[5],90% 以上的大气廓线能够观测到离地面 2 km 高度以下的大气层^[6]。通过减少跟踪误差, COSMIC GPS 反演误差显著减少^[7]。这为天气预

报和全球气候分析提供了一个前所未有的机遇。关于 COSMIC GPS 掩星水汽廓线的精度和数据质量的可靠性已经有一系列的文章对其进行阐述^[6,8]。Ho 等利用 2006 年 7—11 月的 COSMIC 反演的比湿数据和欧洲中尺度天气预报中心 (European Centre for Medium-range Weather Forecasts, ECMWF) 提供的比湿数据进行比较,发现两种数据之间几近零偏差^[6];2011 年, Kishore 等利用热带地区 13 个无线电探空站的比湿廓线对 COSMIC 反演的比湿数据进行了详细的验证,结果验证了在 8 km 高度层以下, COSMIC 的比湿廓线是可靠的^[8]。然而,前人对 COSMIC GPS 掩星反演的比湿廓线进行验证时, COSMIC 比湿廓线与 ECMWF 再分析数据或者与无线电探空站数

* 收稿日期:2016-01-31
基金项目:国家重点基础研究发展计划资助项目(2013CB733302);国家自然科学基金资助项目(41131067,41374036);武汉大学大学生创新创业训练资助项目(S2015777)
作者简介:陈志平(1988—),男,江西抚州人,博士研究生,E-mail: zhpcchen@whu.edu.cn;
罗佳(通信作者),男,副教授,博士,博士生导师,E-mail: jl原因@sgg.whu.edu.cn

据的配对标准都不尽严格。另外,对 COSMIC 比湿廓线验证的时间周期和区域都具有明显的局限性。这将导致对 COSMIC GPS 掩星反演的水汽廓线的质量验证不够严谨。

无线电探空仪(radiosonde)是目前气象业务中实地探测地球水汽廓线最常用的工具。无线电探空数据是数值天气预报(Numeric Weather Prediction, NWP)系统的基本输入资料,长期以来作为标准观测值对各种气象卫星资料进行校正^[9-10]。在理想条件下,用无线电探空观测能够得到较高质量的气象参数数据,温度能够达到 $\pm 0.2\text{ K}$,气压达到 $\pm 0.5\text{ hPa}$,相对湿度达到 $\pm 2\%$ 以上^[3,11-12]。一般情况下,探空站一天早、晚各观测一次,并且全球分布极不均匀。同时,由于探空气球上传感器本身的局限,探空站观测的水汽廓线不可避免地引入了一定程度的误差。而且,探空站在不同区域采用的探空仪和传感器型号不尽相同,探空资料的误差在不同区域并不一致^[13-14]。并且探空站释放的探空气球在上升过程中是随风飘动的,因此探空资料还存在水平漂移误差。一般情况下,探空气球在低对流层的平均漂移距离在几千米左右,在对流层中部的平均漂移距离约为 5 km ,在对流层上部的平均漂移距离约为 20 km ^[15]。同时,高度在 $1\sim 10\text{ km}$ 高度范围内,COSMIC 掩星点的切点的漂移距离约为 102 km ,在 $1\sim 20\text{ km}$ 高度范围内的掩星点切点的水平漂移距离约为 136 km ^[14]。因此,本文在比较无线电探空站与 COSMIC 掩星反演的比湿廓线时采用两者之间的时间观测限值为 1 h ,空间距离差限值采用 100 km 的配对标准。

1 数据和方法

为了消除数据的季节及年际变化对 COSMIC GPS 掩星反演比湿廓线的影响,本文只采用了 2006—2012 年北半球冬季(12 月、1 月、2 月)的 COSMIC GPS 掩星反演的比湿数据与低纬度地区的无线电探空数据及大气成分和气候监测(Monitoring Atmospheric Composition & Climate, MACC)再分析数据进行比较分析。

1.1 COSMIC GPS 掩星观测数据

在中性大气层内(包括对流层和平流层),大气折射率 N 可用多普勒频移和弯曲角反演得到^[2]。由弯曲角反演的大气折射率 N 与温度 $T(\text{K})$ 、大气压 $P(\text{hPa})$ 及水汽压 $e(\text{hPa})$ 有关,公式如下:

$$N = 77.6 \frac{P}{T} + 3.73 \times 10^5 \frac{e}{T^2} \quad (1)$$

由式(1)可以看出,在中性大气层中,GPS 掩星观测资料包含了水汽和温度的信息。通过一维变分同化技术,可以反演出 e 和 T 的廓线。

文中所用的 COSMIC 水汽廓线是由 COSMIC 数据分析与处理中心(COSMIC Data Analysis and Archive Center, CDAAC)提供的“wetprf”数据^①。其中比湿 q 可由式(2)得出^[11]:

$$q = \varepsilon \cdot \frac{e}{p - (1 - \varepsilon)e} \quad (2)$$

式中, $\varepsilon = R_d/R_v$ (≈ 0.622), R_d 、 R_v 分别为干空气及水汽的气体常数。

1.2 IGRA 无线电探空数据

本文利用美国国家气候数据中心(National Climatic Data Center, NCDC)提供的综合全球探空资料(Integrated Global Radiosonde Archive, IGRA)探空站数据来对 COSMIC GPS 掩星数据进行可靠性验证。IGRA 数据是由分布在全球 1500 多个探空站和测风气球在各个年代观测的数据组成,许多测站的数据从 20 世纪 60 年代就开始记录。IGRA 探空站数据质量通过严格的检测,其可靠性也早已被验证^[16-17]。如引言分析,本文采用时间观测限值 1 h ,空间距离观测限值 100 km 配对标准来比较 COSMIC GPS 掩星及 IGRA 探空站所得到的水汽廓线。

1.3 MACC 再分析数据

MACC 再分析数据是 ECMWF 提供的一种监测大气成分和气候的分析数据。MACC 再分析数据提供了 2003—2012 年的比湿数据。MACC 再分析数据的垂直分辨率为 23 层,气压值从 1000 hPa 到 1 hPa 分布,水平分辨率为 $1.125^\circ \times 1.125^\circ$ 。依据 COSMIC GPS 掩星廓线的时间和经纬度,将 MACC 再分析数据内插成与 COSMIC GPS 掩星廓线同时同地发生的廓线,这样可以基本上解决两者之间的配对问题。

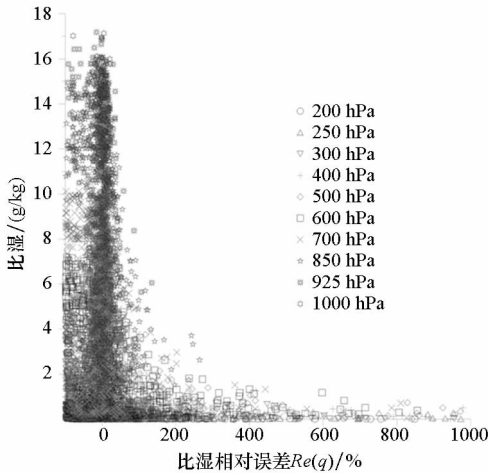
2 结果和讨论

2.1 COSMIC GPS 掩星观测数据与无线电探空数据之间的比较分析

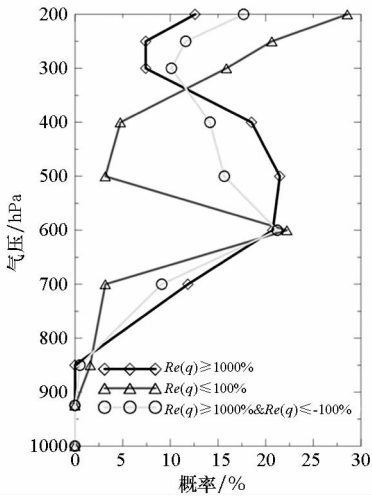
利用低纬度地区($30^\circ\text{S} \sim 30^\circ\text{N}$)384 个探空站来与 COSMIC 反演的比湿廓线进行比较。众所周

① <http://cosmicio.cosmic.ucar.edu/cdaac/index.html>

知,在对流层,随着高度的增加,比湿值逐渐减小。且在对流层上层,比湿值很小。另外,在对流层上层,无线电探空站所观测到的比湿值的精度会降低^[18-19]。本文采用比湿的相对误差 $Re(q)$ 来进行统计比较,如式(3)所示。式(3)中 q_{RS} 、 q_{COSMIC} 分别为探空站及 COSMIC 反演的比湿值。为了避免比湿相对偏差 $Re(q)$ 的极端误差影响比较结果,分析了 $Re(q)$ 的分布情况,由图 1 可知, $Re(q)$ 绝大部分都分布在 $-100\% < Re(q) < 1000\%$ 范围内,且主要分布在 $-100\% < Re(q) < 100\%$ 范围内(见图 1(a)), COSMIC 反演的比湿值比较



(a) COSMIC 水汽廓线相对误差的分布情况
(a) Distribution of the relative deviation of specific humidity profiles



(b) COSMIC 水汽廓线极端值的概率分布图
(b) Distribution of extreme error of COSMIC specific humidity profiles

知,在 600 hPa 高度层,出现极端值的概率较大,负极端值个数为 63,正极端值个数为 135。在统计的对流层不同高度层范围内,只有很少比湿廓线被排除。故而,本文采用 $Re(q) \geq +1000\%$ 及 $Re(q) \leq -100\%$ 来排除极端比湿廓线是可靠的,此结果与王伯睿等的研究结果类似^[14]。

$$Re(q) = \frac{q_{RS} - q_{COSMIC}}{q_{COSMIC}} \times 100\% \tag{3}$$

由前文分析可知,本文采用时间差限值 1 h,空间距离差限值 100 km 来对 COSMIC 反演的比湿廓线及探空站观测的比湿廓线进行配对。总共约有 1120 对廓线,且在 1000 hPa ~ 200 hPa 高度层中,约有 8252 对配对的比湿点值。其中有 8054 (97.6%) 个比湿点值分布在 $-100\% < Re(q) < 1000\%$ 区间上。

图 2 描述了配对好的探空数据与 COSMIC 反演的比湿廓线之间的统计结果比较。由图 2 可知,平均比湿相对误差在统计的各高度层上都小于 20%,比湿相对误差的标准差也小于 20%。在热带地区,探空比湿廓线值要普遍大于 COSMIC 反演的比湿值。统计的所有高度层的平均比湿相对误差为 8%,平均比湿相对误差的标准差为 16%。探空比湿数据相对于 COSMIC 反演的比湿数据的平均相对误差在不同高度是不一样的。在 700 hPa 到 400 hPa 高度层,相对误差呈现正值 (<20%),600 hPa ~ 500 hPa 高度层出现最大的正向相对误差,这主要是因为这段高度层中水汽含量对 COSMIC 反演有很大的影响。在 850 hPa 以下高度层及对流层上层,无线电探空站的值比 COSMIC 反演的比湿值小,两者之间的负相对误

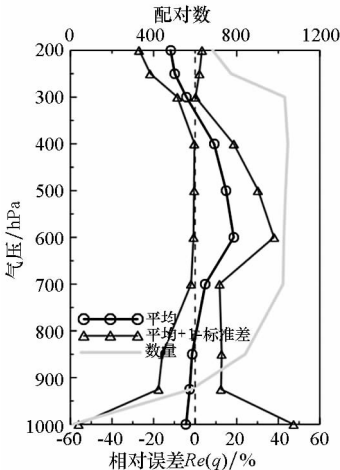


图 2 无线电探空站与 COSMIC 掩星反演的比湿廓线之间的比较统计结果

Fig.2 Comparison of specific humidity profiles between collocated COSMIC and radiosonde soundings

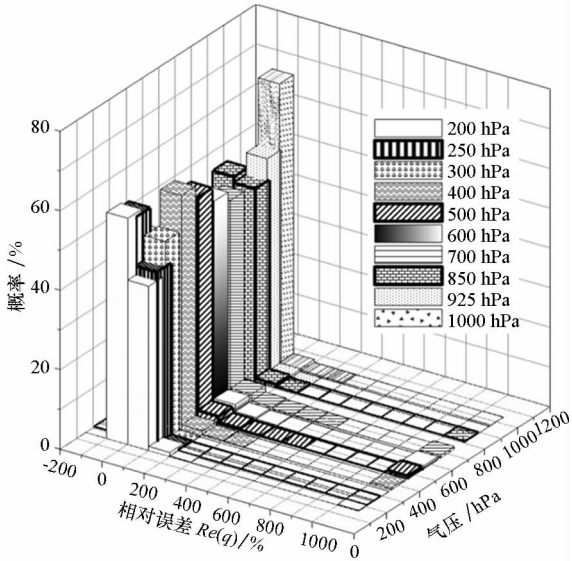
图 1 COSMIC 反演的水汽廓线相对于无线电探空数据的相对误差的分布
Fig.1 Distribution of the relative error of COSMIC specific humidity relative to radiosonde dataset
小,则 $Re(q)$ 出现极端值概率更大。由图 1(b)可

差分别小于 5.5% 及 12%。这主要是由于在低对流层, COSMIC 掩星廓线能达到地球表面的廓线数量较少。在对流层上层, 无线电探空站的比湿值出现缺失, 这在图 2 中探空站及 COSMIC 掩星廓线的配对数量中也能反映出来。

2.2 COSMIC GPS 掩星观测数据与 MACC 再分析数据之间的比较分析

为了更好地分析水汽对 COSMIC GPS 掩星的影响, 本文给出了 COSMIC 反演的水汽廓线与 MACC 再分析数据之间的统计结果比较。与 2.1 部分一样, 这里采用了相同的统计方法。不过鉴于低纬度地区水汽含量过高, 这里不仅对低纬度地区(30°S ~ 30°N)进行了统计分析, 同时对全球范围内的 COSMIC GPS 掩星比湿廓线也做了统计分析。尽管 MACC 比湿数据相对于 COSMIC GPS 掩星反演的比湿廓线的相对误差分布与图 1 类似, 但由于数据量过大, 再用图 1(a) 这种形式并不能很好地描述 MACC 再分析数据相对于 COSMIC GPS 掩星反演的比湿廓线的相对误差分布, 其分布如图 3 所示。图 3 中, 横轴 $-200\% < Re(q) < -100\%$ 表示 $Re(q) < -100\%$ 的分布, 横轴 $1000\% < Re(q) < 1100\%$ 代表 $Re(q) > 1000\%$ 的分布。

由图 3 可以看出, 在 2006 年 12 月至 2012 年 12 月这段时期的冬季, 约有 288 663 (883 041) 对 MACC&COSMIC 比湿廓线分布于低纬度地区(全球范围)。在低纬度地区(全球范围), 有 95.6% (97.6%) 的比湿相对误差在 $-100\% < Re(q) <$

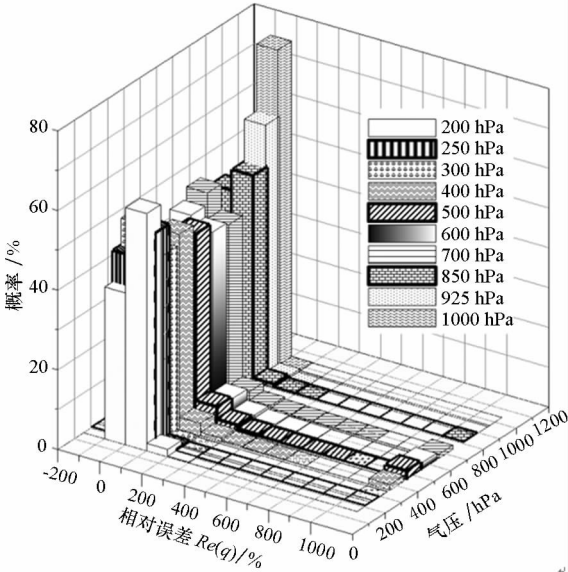


(b) 全球比湿廓线的相对误差的概率分布
(b) Distribution of relative deviations of specific humidity for globe

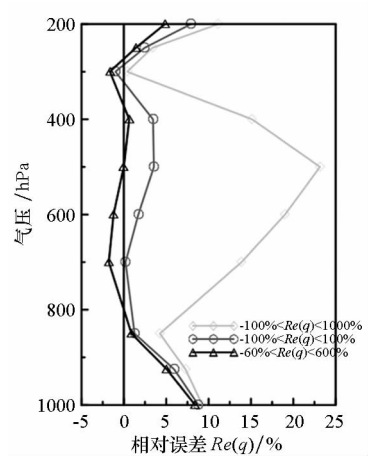
图 3 MACC 再分析数据相对于 COSMIC 比湿廓线的相对误差的概率分布
Fig. 3 Probability distribution of relative deviation between MACC-reanalysis and COMSIC specific humidity profiles

100% 范围内。在 $Re(q) \leq -100\%$ 范围内, 在统计的高度层, 几乎没有比湿廓线。全球范围内, 99% 以上的比湿相对误差在 $-100\% < Re(q) < 1000\%$ 范围内。为了使水汽对 COSMIC 反演的结果能够更好地量化, 采用比湿相对误差位于不同区间内的取值, 来分析 MACC 再分析比湿数据相对于 COSMIC GPS 掩星反演的比湿数据的质量变化。

图 4 描述了 $Re(q)$ 不同范围内, MACC 再分析比湿数据相对于 COSMIC GPS 掩星反演的比湿数据的平均相对误差随高度的变化。在全球范围内 $-100\% < Re(q) < 1000\%$, 除了在 250 hPa 高度层(平均相对误差为 -0.645%) 以外, MACC 再分析数据比 COSMIC GPS 掩星比湿值小。在 700 hPa 到 400 hPa 高度层范围, MACC 再分析数据相对于 COSMIC GPS 掩星反演的比湿数据的正平均相对误差随着 $Re(q)$ 范围的减小急剧减小。在低纬度地区也有相似的情况。在低纬度地区, 相比于全球范围, MACC 再分析数据显示一个更大的平均相对误差, 这可能是因为 700 hPa 到 400 hPa 高度范围, 低纬度地区的水汽含量更加充沛, 这对 GPS 掩星获取大气廓线有很大的影响。在 850 hPa 以下高度层, 由于大气水汽含量的影响以及 COSMIC 掩星廓线很少能到达 850 hPa 以

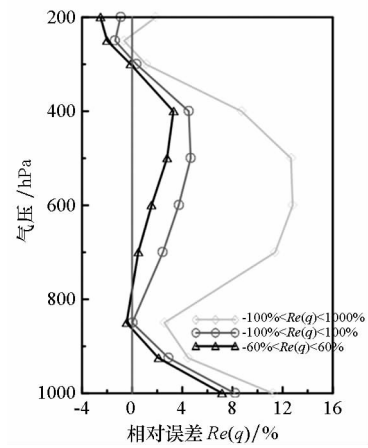


(a) 低纬度地区比湿廓线的相对误差的概率分布
(a) Distribution of relative deviations of specific humidity for low latitude



(a) 低纬度地区 COSMIC 比湿廓线的相对误差随 $Re(q)$ 值不同区间的变化

(a) Relative deviation between the specific humidity profiles from the COSMIC observations over low latitude areas under different relative deviation ranges



(b) 全球范围 COSMIC 比湿廓线的相对误差随 $Re(q)$ 值不同区间的变化

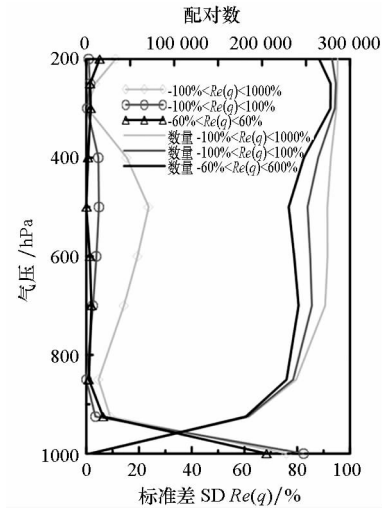
(b) Relative deviation between the specific humidity profiles from the COSMIC observations over globe under different $Re(q)$ ranges

图4 MACC 数据相对于 COSMIC 比湿廓线的相对误差随 $Re(q)$ 值不同区间的变化图

Fig.4 Relative deviation between the specific humidity profiles from the COSMIC observations and MACC – Reanalysis under different $Re(q)$ ranges

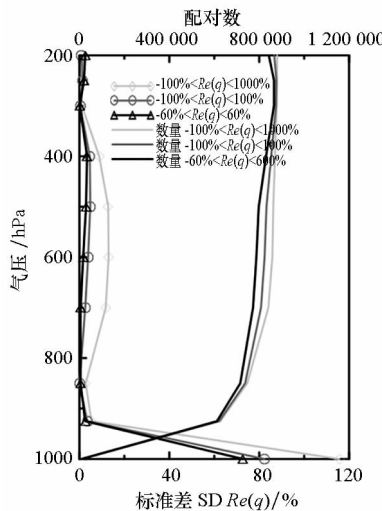
下,故而导致 MACC 再分析比湿数据相对于 COSMIC 比湿的相对误差随着高度的降低而逐渐增大。

图5描述了 $Re(q)$ 不同区间范围内,MACC 再分析数据相对于 COSMIC GPS 掩星反演的比湿观测值的标准差随高度的变化。由图5可以看出,随着 $Re(q)$ 区间范围的急剧缩小,掩星点值的数量变化却很小。但 MACC 再分析数据相对于 COSMIC GPS 掩星反演的比湿平均相对误差及标



(a) 低纬度地区 COSMIC 比湿廓线的标准偏差随 $Re(q)$ 值不同区间的变化图

(a) Standard deviation between the specific humidity profiles from COSMIC observations over low latitude under different $Re(q)$ ranges



(b) 全球范围内 COSMIC 比湿廓线的标准偏差随 $Re(q)$ 值不同区间的变化

(b) Standard deviation between the specific humidity profiles from COSMIC observations over globe under different $Re(q)$ ranges

图5 MACC 再分析比湿数据相对于 COSMIC 比湿廓线的标准偏差随 $Re(q)$ 值不同区间的变化

Fig.5 Standard deviation between the specific humidity profiles from COSMIC observations and MACC – Reanalysis under different $Re(q)$ ranges. And curves without symbols show the number of data points

准差在统计的高度层区间变化很大,尤其是当排除极端相对误差范围区间由 $Re(q) \geq 1000\%$ & $Re(q) \leq -100\%$ 缩小至 $Re(q) \geq 100\%$ & $Re(q) \leq -100\%$ 时,在 700 hPa 到 400 hPa 高度范围内变化尤其明显。针对全球范围, $Re(q)$ 限制范围为

$-100\% < Re(q) < 1000\%$ ($-100\% < Re(q) < 100\%$; $-60\% < Re(q) < 60\%$) 时,统计高度层的平均相对误差为 6.62% (2.45%; 1.23%) (见图 4(b)), 平均标准差为 17.37% (10.44%; 8.90%) (见图 5(b))。对于低纬度地区,统计高度层的平均相对误差为 10.69%, 3.43%, 1.66% (见图 4(a)); 平均标准差依次为 17.73%, 10.11%, 8.79% (见图 5(a))。这些都充分说明了水汽对 COSMIC GPS 掩星反演的影响很大。

3 结论

1) 无论是与低纬度地区的探空站还是 ECMWF 提供的 MACC 再分析比湿数据比较,总体上,COSMIC GPS 掩星反演的比湿廓线与它们之间的差异性并不大,有很好的一致性。

2) 水汽对 COSMIC GPS 掩星反演有很大的影响。在对流层中部(700 hPa ~ 400 hPa)以及对流层低层(850 hPa 以下)影响尤其明显,在850 hPa 高度以下,COSMIC GPS 掩星反演的比湿廓线质量与精度随着高度的减小明显下降。在对流层中部及上层,出现极端误差值的概率明显变大。

3) 无线电探空站比湿数据在对流层上层的质量及精度明显不足,同时 COSMIC GPS 掩星反演的比湿数据在低对流层(850 hPa 高度以下)的精度也存在着问题。

参考文献 (References)

[1] Anthes R A, Rocken C, Kuo Y H. Applications of COSMIC to meteorology and climate[J]. Terrestrial Atmospheric and Oceanic Sciences, 2000, 11(1): 115 - 156.

[2] Kursinski E R, Hajj G A, Schofield J T, et al. Observing Earth's atmosphere with radio occultation measurements using the global positioning system[J]. Journal of Geophysical Research: Atmospheres, 1997, 102(D19): 23429 - 23465.

[3] Ware R, Rocken C, Solheim F, et al. GPS sounding of the atmosphere from low Earth orbit: preliminary results[J]. Bulletin of the American Meteorological Society, 1996, 77(1): 19 - 40.

[4] Rocken C, Ying-Hwa K, Schreiner W S, et al. COSMIC system description[J]. Terrestrial Atmospheric and Oceanic Sciences, 2000, 11(1): 21 - 52.

[5] Anthes R A, Ector D, Hunt D C, et al. The COSMIC/FORMOSAT-3 mission: early results[J]. Bulletin of the American Meteorological Society, 2008, 89(3): 313 - 333.

[6] Ho S P, Zhou X J, Kuo Y H, et al. Global evaluation of

radiosonde water vapor systematic biases using GPS radio occultation from COSMIC and ECMWF analysis[J]. Remote Sensing, 2010, 2(5): 1320 - 1330.

[7] Sokolovskiy S V, Rocken C, Lenschow D H, et al. Observing the moist troposphere with radio occultation signals from COSMIC[J]. Geophysical Research Letters, 2007, 34(18): 266 - 278.

[8] Kishore P, Ratnam M V, Namboothiri S P, et al. Global (50 S - 50 N) distribution of water vapor observed by COSMIC GPS RO: comparison with GPS radiosonde, NCEP, ERA-Interim, and JRA - 25 reanalysis data sets[J]. Journal of Atmospheric and Solar-Terrestrial Physics, 2011, 73(13): 1849 - 1860.

[9] 徐晓华, 张克非, 罗佳. GPS 掩星与探空观测统计比较中配对标准的比较研究[J]. 武汉大学学报: 信息科学版, 2009, 34(11): 1332 - 1335.

XU Xiaohua, ZHANG Kefei, LUO Jia. Research on the collocation criteria in the statistical comparisons of GPS radio occultation and radiosonde observations[J]. Geomatics and Information Science of Wuhan University, 2009, 34(11): 1332 - 1335. (in Chinese)

[10] Xu X, Luo J, Shi C. Comparison of COSMIC radio occultation refractivity profiles with radiosonde measurements[J]. Advances in Atmospheric Sciences, 2009, 26(6): 1137 - 1145.

[11] Peixoto J P, Oort A H. The climatology of relative humidity in the atmosphere[J]. Journal of Climate, 1996, 9(12): 3443 - 3463.

[12] Wickert J. Comparison of vertical refractivity and temperature profiles from CHAMP with radiosonde measurements[M]. Geoforschungszentrum Potsdam, 2005.

[13] Sun B, Reale A, Seidel D J, et al. Comparing radiosonde and COSMIC atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collocation on comparison statistics[J]. Journal of Geophysical Research: Atmospheres, 2010, 115(D23): 6696 - 6705.

[14] Wang B R, Liu X Y, Wang J K. Assessment of COSMIC radio occultation retrieval product using global radiosonde data[J]. Atmospheric Measurement Techniques, 2013, 5(6): 8405 - 8434.

[15] Seidel D J, Sun B, Petty M, et al. Global radiosonde balloon drift statistics[J]. Journal of Geophysical Research: Atmospheres, 2011, 116(D7): D07102.

[16] Eskridge R E, Alduchov O A, Chernykh I V, et al. A comprehensive aerological reference data set (CARDS): rough and systematic errors[J]. Bulletin of the American Meteorological Society, 1995, 76(10): 1759 - 1775.

[17] Durre I, Vose R S, Wuertz D B. Overview of the integrated global radiosonde archive[J]. Journal of Climate, 2006, 19(1): 53 - 68.

[18] Elliott W P, Gaffen D J. On the utility of radiosonde humidity archives for climate studies[J]. Bulletin of the American Meteorological Society, 1991, 72(10): 1507 - 1520.

[19] Soden B J, Lanzante J R. An assessment of satellite and radiosonde climatologies of upper-tropospheric water vapor[J]. Journal of Climate, 1996, 9(6): 1235 - 1250.