

一种卫星对地观测任务完成概率的估计模型*

李志猛¹, 谈群², 刘刚¹, 郭少阳¹

(1. 国防科技大学 信息系统工程重点实验室, 湖南长沙 410073; 2. 95980 部队, 湖北襄樊 441000)

摘要:针对及时响应用户提出的观测需求和高效实施卫星任务规划调度的需要,提出快速估计卫星对地观测任务完成可能性的现实问题,利用 Logistic 回归方法建立了一种卫星对地观测任务完成概率的估计模型。分析影响侦察任务完成概率的相关因素,包括侦察任务自身属性、资源约束、任务之间竞争关系等;从影响因素中提炼出刻画影响因素的模型变量,给出其量化方法,并初步构建了 Logistic 回归分析模型;基于卫星成像侦察任务规划系统开展实验,获取自变量与规划结果方面的样本数据;采用 Logistic 回归分析方法对样本数据进行分析,确定模型中的变量、参数及函数形式。结论表明,模型具有很好的统计特性。

关键词:卫星对地观测;任务完成概率;Logistic 回归模型

中图分类号:TN971 文献标志码:A 文章编号:1001-2486(2013)06-0185-06

An enforceability probability estimation model for satellite observing tasks

LI Zhimeng¹, TAN Qun², LIU Gang¹, GUO Shaoyang¹

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

2. Unit 95980, Xiangfan 441000, China)

Abstract: The enforceability probability of satellite observation requires timely and accurate estimation to provide the fundamental basis for satellite task planning and scheduling. An estimation model of the enforceability probability of satellite earth observation is established based on the Logistic regression method. Firstly, the influence factors of observation enforceability probability were analyzed, including the property of observation task, resources limitation, competitive relations among tasks, etc. Secondly, independent variables were extracted and their quantitative methods were given, then the Logistic regression model was established. Thirdly, experiments were implemented on the planning and scheduling system of satellite imaging observation to obtain the sample data of independent variables and planning results. Moreover, the sample data was analyzed by using the Logistic regression method, and then the variables and parameters of the model and the function forms were all determined. Lastly, the effectiveness of the model was validated.

Key words: satellite observing task; enforceability probability; logistic regression method

卫星对地观测以其独特的时空范围、精度、可靠性等不可替代的优势,在现代社会中,特别是军事领域扮演着重要的角色。由于卫星资源的稀缺性以及卫星平台的时空不连续特性,动态变化的用户需求往往无法得到完全满足。目前,用户需求提交后,还需要经过卫星规划管理部门的评估与调度后才能得到响应——实际执行或无法满足,这一过程往往需要很长时间,很难满足高时效任务的要求。如果能够建立侦察任务完成概率估计模型,准确地预估出观测任务完成的可能性,对于总体掌控与有效管理用户需求,更高效实施卫星任务规划调度,满足更多更重要任务的要求具有重要意义。

当前关于卫星对地观测任务完成可能性估计

的研究较少,文献[1]提出了一种基于任务流模型的任务完成概率估计模型,该模型针对给定的任务集及卫星资源,估计任务的整体完成率。文献[2]在此基础上提出了估算每个任务完成概率的模型,这两种模型都是经验模型,可验证性较差。在社会学、心理学、经济学等领域中,利用 Logistic 回归模型对“二分因变量”进行回归分析是较为普遍的多元量化分析方法^[3],具有逻辑清晰、精度高、适应性强等优点,尤其适用于分析多类复杂因素对事件发生概率的影响。本文以实验数据为基础,采用 Logistic 回归分析方法,建立了一种卫星对地观测任务完成概率估计模型,并通过统计检验方法验证了模型的有效性。

本文提出的估计模型可为卫星规划管理部门

* 收稿日期:2013-04-01

基金项目:国家部委资助项目

作者简介:李志猛(1978—),男,安徽宿州人,讲师,博士,E-mail:zml@nudt.edu.cn

审核观测任务需求、提出合理的需求满足方案提供定量依据,同时,也有利于及时向用户反馈其任务需求的预期满足程度,有利于作战方案制定及作战准备工作开展。

1 问题分析

卫星对地观测任务是指在特定时间、空间条件下,以卫星观测为手段,为满足用户特定信息需求而进行的对地观测及相关活动。一般过程是用户先提出对地观测任务申请,任务管理中心汇总任务申请后进行一定的处理,然后任务规划系统进行规划调度并生成上星指令,经地面站上传至卫星。卫星对目标进行观测并下载数据,地面信息处理系统经技术处理和分析判读,从中获取有价值的情报,分发给用户。在这个过程中,最难估计的环节就是任务规划,卫星完成对地观测任务本身复杂的约束条件及任务间可能发生的冲突都会增加估计的难度。卫星对地观测任务的完成与否理论上包括两个方面,一是任务提交到任务规划器后成功争取到观测资源的概率,二是成功规划后卫星实际运行成功实施观测的概率。由于后者主要取决于卫星系统的可靠性与观测环境的具体情况,本文不作研究。即本文中卫星对地观测任务的完成概率估计不考虑争取到资源但任务仍然没有成功执行的情况。

卫星对地观测任务完成的可能性主要受三个方面的因素影响:一是观测任务本身的属性,如任务优先级、有效时间、质量要求等;二是卫星观测资源,包括资源的能力及运行状态等;三是与其他观测任务可能发生的冲突。

第一,卫星对地观测任务自身的属性要求对任务完成概率影响是显然的。任务规划研究中目标函数常取完成任务的优先级之和最大^[5],由此,优先级越高的任务,完成概率越大。有效时间越长的任务,往往会有更多的时间窗口,因此相应特征的任务完成概率也就越大。在资源水平较低时,一些时效性要求高的任务完成率非常低,即使这些任务非常重要。任务的观测对象和质量要求也会对任务完成概率造成影响,质量要求越高,有能力完成该任务的资源就越少,任务完成概率就越低。

第二,卫星对地观测资源越多、能力越强,任务完成概率也就越高。卫星在任务有效时间内过顶观测对象的次数直接影响任务完成概率,这种影响可以通过任务的时间窗口个数反映出来,任务的时间窗口个数越多,完成概率越高,如果侦察任务没有时间窗口,则侦察任务完成概率为 0。

第三,卫星对地观测任务之间的冲突会降低任务完成概率。冲突可能来自于多个方面,如时间、空间上都有重叠的任务竞争同一个卫星资源时可能发生时间窗口冲突,大量任务竞争一个卫星资源时,可能超出卫星的能量或存储限制,这些冲突导致一些有时间窗口的任务无法获得资源。可以通过任务的时空分布来估计任务之间发生冲突的可能性。一般来说,任务在时间上、空间上越密集,冲突的可能性就越大。

各个影响因素对任务完成概率的影响往往不是互相独立的,如任务的质量要求与观测资源的能力共同决定了满足任务要求的观测资源数量,观测资源的能力对任务之间的冲突概率也有影响。建立预估模型时需要注意影响因素之间的相关性。

2 模型构建

2.1 模型自变量及其量化方法

根据上面影响因素分析的结果,选取能够充分反映这些影响因素的指标作为任务完成概率估计模型的自变量。

任务的自身属性中,优先级与有效时间长度会影响完成概率,且与其他影响因素关系不大,容易获取,可以选取优先级作为模型的量化指标之一。任务的质量要求影响了有能力完成任务的卫星资源的数量,在计算时间窗口时要充分反映这一点。

卫星对地观测资源的影响最终表现为时间窗口数量,通过 STK 等专业软件可以准确计算每个任务的时间窗口,但有时由于资源、任务数量太大,准确计算太过繁琐。本文用较易获取的指标(如日平均时间窗口密度)来替代,为了便于比较,以每小时内的平均时间窗口个数作为资源度量指标。

卫星对地观测任务之间发生冲突的可能性与任务在时间上、空间上的密度成正比。空间维任务密度较容易计算,较粗略但简便的方法是将整个区域划分为若干个小区域,统计每个小区域内的任务量,然后用此任务量去除每个小区域的面积即可。时间维任务密度计算较复杂,文献[6]提出了一种卫星对地观测任务流的时间密度函数,其中自变量为具体时间,函数值表示相应时刻待完成的任务总量。该方法在这里不适用,因为如果任务的平均有效时间越长,按此计算,时间密度函数值就越大,而事实并不总是如此。本文提出一种新的时间维任务密度函数,基本思想是综

合考虑优先级和任务有效时间。对一个任务来说,若其有效时间内其他高优先级的任务越多,说明该任务竞争到资源的概率越小,设定其在单位时间内对任务流强度函数贡献就越小。

设有任务 $task_1, task_2, \dots, task_n$, 有效时间区间及其长度分别记为 $T_i, L(T_i)$, 优先级分别记为 λ_i , 则任务数量在时间轴上分布的度量函数如式(1)所示。

$$f(t) = \sum_{i=1}^n \Phi(t, T_i) \cdot \lambda_i / L(T_i) \quad (1)$$

其中:

$$\Phi(t, T_i) = \begin{cases} 1 & t \in T_i \\ 0 & t \notin T_i \end{cases}$$

任务 $task_i$ 的时间维任务密度函数由式(2)给出。

$$td_i = \frac{\int_{T_i} f(t) dt}{L(T_i)} \quad (2)$$

如上所述,得到可能用于卫星对地观测任务完成概率建模的自变量如表 1 所示。

表 1 模型自变量列表

Tab. 1 Variables list of the estimation model

指标名称	符号(简记)	单位	量化方法	影响类型
优先级	$pri(r)$	无	任务自身属性	正相关
有效时间长度	$long(l)$	h	任务有效时间结束时刻减去开始时刻	正相关
有效时间内时间窗口密度	$task_windows(tw)$	个/h	任务有效时间内时间窗口个数除以有效时间长度	正相关
时间维任务密度	$time_density(td)$	个/h	如式(2)所示	负相关
空间维任务密度	$space_density(sd)$	个/km ²	区域任务数量/区域面积	负相关

2.2 估计模型的函数形式

根据上面的分析,卫星对地观测任务完成概率与多类影响因素相关,这些影响因素和完成概率之间的关系一般不是线性关系,且对完成概率影响的大小无法事前确知。这些特征正是统计分析中应用 Logistic 回归模型的一般要求,本文采用 Logistic 回归模型形式进行研究,并基于实验数据对模型进行检验和修正,最终说明 Logistic 回归模型的可用性。

设卫星对地观测任务完成发生概率为 cp , 一般地,可用以下 Logistic 函数来表示 cp 与自变量的关系:

$$\ln \frac{cp}{1 - cp} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_k \cdot x_k \quad (3)$$

其中 $cp/(1 - cp)$ 是事件发生概率与不发生概率的比值,简称为事件的发生比, x_k 为模型自变量, b_k 是各自变量的影响系数。当等式右边从 $-\infty \rightarrow +\infty$ 变化时, cp 在 $[0, 1]$ 区间上单调递增,等式右边绝对值越大时, cp 变化越平缓。

使用表 1 中各类影响因素的命名,并对式(3)进行变换,得

$$cp = \frac{\exp(b_0 + b_1 \cdot r + b_2 \cdot l + b_3 \cdot tw + b_4 \cdot td + b_5 \cdot sd)}{1 + \exp(b_0 + b_1 \cdot r + b_2 \cdot l + b_3 \cdot tw + b_4 \cdot td + b_5 \cdot sd)} \quad (4)$$

将式(4)作为一个初始模型进行分析,后期通过模型估计再考虑增减变量或完善函数形式。

3 实验与数据

3.1 实验设计

根据实验设计的随机化原则、重复原则、对照原则和分组原则^[8],本文随机产生卫星对地观测任务。卫星对地观测资源则设定几种水平,在各种组合下进行多次实验,获取充足的样本。

卫星对地观测资源主要根据卫星数量设为 10、20、50 三种水平,分别在 STK 中建立场景,供任务规划系统调用。根据卫星数量,每次产生一组卫星对地观测任务,输入任务规划系统进行规划。在卫星数量不同时,产生的任务数量也不同,为了使各个区组之间的样本数量保持均衡,每次产生任务数量少的实验重复更多次数,最终的实验方案如表 2 所示。

表 2 实验方案
Tab.2 Experiment design

方案编号	卫星数量	任务数量	时间分布	有效时间长度	实验次数	样本数量
1	10	90	泊松分布	0.5 ~ 30	5	450
2	10	90	非平稳泊松分布	0.5 ~ 30	5	450
3	10	180	泊松分布	0.5 ~ 30	2	360
4	10	180	非平稳泊松分布	0.5 ~ 30	3	540
5	20	180	泊松分布	0.5 ~ 20	2	360
6	20	180	非平稳泊松分布	0.5 ~ 20	3	450
7	20	225	泊松分布	0.5 ~ 20	2	450
8	20	225	非平稳泊松分布	0.5 ~ 20	2	450
9	50	225	泊松分布	0.5 ~ 15	2	450
10	50	225	非平稳泊松分布	0.5 ~ 15	2	450
11	50	450	泊松分布	0.5 ~ 15	1	450
12	50	450	非平稳泊松分布	0.5 ~ 15	1	450

3.2 数据采集

本文在卫星对地观测任务规划系统上开展实验,分为任务生成、任务规划、数据采集三个步骤。首先,根据实验设计的结果每次随机产生一组任务写入数据库,向卫星对地观测任务规划系统发送任务规划指令后,任务规划系统读取任务进行规划,最后将指标值及规划结果写入指定的数据文件。

每次数据采集的内容包括任务 ID、优先级、有效时间长度、任务有效时间内时间窗口密度、日平均时间窗口密度、时间维任务密度、空间维任务密度以及任务规划结果。完成所有实验后,一共获得 5400 条数据,基于这些数据,对上面式(4)中模型进行回归分析。

4 模型估计与检验

4.1 变量估计

采用统计软件 SPSS 分析处理实验数据。经过多次增减变量,反复比较,最终以规划结果为因变量,以优先级、有效时间长度、有效时间内时间窗口密度、时间维任务密度、空间维任务密度为协变量,选择变量进入分析的方式为“强迫引入法”,在选项中选中“Hosmer-Lemeshow goodness-of-fit”与“Correlations of estimates”,得到分析结果如表 3 所示。

表格第一列显示变量名称,第二列表示每个

自变量在模型中的系数,第四列是系数的 Wald 检验值,Wald 检验是 Logistic 回归中常用的回归系数显著性检验方法,适用于回归系数绝对值不太大的情况,第六列是 Wald 检验的显著性概率,当 Sig. 小于给定的显著性水平时(一般取 0.05,也有学者认为应该取 0.25,避免遗漏重要变量),认为自变量对因变量的影响是显著的。

表 3 变量估计的分析结果
Tab.3 The analysis result for variable estimation

变量	系数	标准差	Wald 检验值	自由度	显著性概率	风险比
<i>r</i>	0.110	0.011	97.019	1	0	1.116
<i>l</i>	0.074	0.004	292.027	1	0	1.076
<i>tw</i>	1.395	0.116	144.040	1	0	4.036
<i>td</i>	-0.026	0.001	301.057	1	0	0.975
<i>sd</i>	-1.322	0.155	72.733	1	0	0.266
常数项	-0.419	0.158	7.077	1	0.008	0.658

从表 3 中可以看到,回归系数的绝对值都不大,因此 Wald 检验是适用的,Sig. 列的结果表明这 5 个自变量的影响都是显著的。此外,各自变量的系数差别很大,最重要的两个自变量 *l* 与 *td* 的系数都很小,*tw* 的系数反而是它们的几十倍,这是由于自变量的取值范围都不相同,并非系数小的变量影响就小。此外,对自变量之间的相关性作了一些处理。通过变量估计,得到各变量的系数,如下

$$cp = \frac{\exp(-0.419 + 0.11r + 0.074l + 1.395tw - 0.026td - 1.322sd)}{1 + \exp(-0.419 + 0.11r + 0.074l + 1.395tw - 0.026td - 1.322sd)} \quad (5)$$

4.2 模型的函数形式估计

在 Logistic 模型中,除了用 Logistic 函数表示

模型因变量与事件概率之间的关系,还要考虑因变量与自变量之间的函数形式。下面考虑存在非

线性与非加性关系的情况。

首先,分析非线性的情况。在模型(5)中有5个自变量,分别加入每个自变量的二次项进行 Logistic 回归分析。这里添加自变量之间的乘积形式,将每个自变量的数据复制一列,重新命名,如 r_2 ,与原来的自变量相乘进入模型,如 $r * r_2$,这样可以模拟每个自变量的二次项。每次添加一个自变量的二次项,得到模型的 χ^2 统计量如表4所示。

表4 增加二次项的 χ^2 统计量
Tab.4 χ^2 statistic with quadric items

增加的二次项	模型的 χ^2 统计量
无	880.392
r by r_2	906.868
l by l_2	891.095
tw by tw_2	893.363
td by td_2	1113.265
sd by sd_2	881.562

从表中的结果可以看到,增加时间维任务密

$$cp = \frac{\exp(-0.13 + 0.38r + 0.07l + 1.30tw - 0.09td - 1.35sd + 0.001td^2 - 0.02r^2 + 0.03l \cdot tw)}{1 + \exp(-0.13 + 0.38r + 0.07l + 1.30tw - 0.09td - 1.35sd + 0.001td^2 - 0.02r^2 + 0.03l \cdot tw)} \quad (6)$$

4.3 模型检验

模型的评价是个复杂的问题,当前没有一个简单而令人满意的标准来衡量 Logistic 回归模型是否足够好,研究人员采用了不同的检验方法来评价模型的优劣。

其中 Hosmer-Lemeshow (HL) 检验是 Hosmer 等在 1989 年提出的一种对 Logistic 回归模型拟合优度的检验方法^[9],当 HL 指标统计不显著时,认为模型拟合优度很好,HL 指标是评价 Logistic 回归模型拟合优度方面应用最广泛的指标。模型系数综合检验中, χ^2 检验是零假设模型与所设模型之间在 $-2LL$ 上的差距,是关于自变量是否与所

$$cp = \frac{\exp(0.332 + 0.115r + 0.079l + 1.543tw - 0.087td - 1.334sd + 0.001td^2)}{1 + \exp(0.332 + 0.115r + 0.079l + 1.543tw - 0.087td - 1.334sd + 0.001td^2)} \quad (7)$$

使用类 R^2 指标进行分析,模型中有两个类 R^2 指标,有人认为该指标能够反映变异中被解释的比例,但是也有人认为不应该将类 R^2 看作模型所解释的变动的比例,因为似然对数根本不是一个可解释的量。他们认为类 R^2 是估计预测准确性的粗略近似,在自变量与因变量完全无关时,类 R^2 值趋于 0,当拟合模型能够完美预测时,类 R^2 值趋于 1 但不等于 1^[10]。当前文献中并没有解释类 R^2 值多大时,模型足够好,许多案例中类 R^2 值都小于 0.5。

文献[11]提出通过分类正确率判断模型是

度的二次项(td^2),模型的 χ^2 统计量提高最显著,在分析结果中,模型的其它指标也有明显提高。 td 与 td^2 的系数分别为 -0.087 与 0.001 ,Wald 检验都是显著的。由于 td 取值范围比较大,因此系数比较小, td 的系数小于 0,说明随着 td 的增加,侦察任务完成概率减小, td^2 大于 0,说明侦察任务完成概率减小的趋势越来越平缓,即随着 td 增加,它对完成概率的影响逐渐变小。经过类似的分析, r 的二次项也被添加到模型中。

其次,分析非加性的情况。 l 与 tw , r 与 td 两组自变量之间有可能存在互动作用,分别加入模型中,进行 Logistic 回归分析得到模型的 χ^2 检验。增加 $r \cdot td$ 得到的 χ^2 统计量为 1143.683, $L. R. = 1.949$,自由度 $d. f. = 8 - 7 = 1$,结果不显著,增加 $l \cdot tw$ 的结果为 1146.813, $L. R.$ 检验显著,因此互动作用项 $l \cdot tw$ 添加到模型中。

经过函数形式估计,得到新的模型如式(6)所示。

研究事件的对数发生比线性相关的检验。

模型(6)的 HL 检验表明自由度为 8,按照 0.05 的显著水平, χ^2 大于 15.5 时是显著的,得到 HL 检验与 χ^2 检验都是非常显著的。

进一步,自变量的 Wald 检验有一部分不显著,8 个自变量 Wald 检验不显著的次数分别为 2、0、1、0、1、1、5、9,这说明前面 6 个自变量对因变量的影响是稳定显著的,而后两个自变量的影响是不稳定的,应该从模型中剔除,去掉这两项后重新分析发现,模型的各项检验指标下降很小,得到新的模型如(7)所示:

否有用,其标准是模型的分类正确率要显著高于一个基准水平,基准水平是指:在一个容量为 n 的样本中,设有 n_1 个观测来自 1 组,有 n_2 个观测来自 0 组, $\max(n_1/n, n_2/n)$ 就是分类正确率的基准水平。按照该方法,本文 4779 个观测中有 2758 个来自于 1 组,基准水平为 57.7%,模型分类正确率为 73.4%,显著高于基准水平。

根据以上分析,任务完成概率模型(7)在 HL 检验、 χ^2 检验、分类正确率等方面都表现良好,可以认为模型是有效的。

5 结论

本文提出了一个卫星对地观测任务完成概率的估计模型,可以在不完全信息情况下快速对任务是否能够完成进行估计,并通过多种方法对模型进行了检验。结果显示,模型具有良好的性能,也存在一些不足有待进一步研究。

(1)实验设计需要考虑更多的现实情况,例如,各优先级任务所占比例应该有一定规律。针对时敏目标的观测任务有效时间长度非常短,往往具有很高的优先级。全面分析卫星对地观测需求后进行实验设计,得到的模型会更符合实际。

(2)实验是在现有的任务规划系统上开展的,对侦察任务进行了一定的简化,如将侦察任务的重要性离散化为优先级,不考虑侦察任务的不确定性及时间偏好等属性,这对估计侦察任务的完成概率及效果有一定影响,下一步可更为细致考虑相关因素。

参考文献 (References)

[1] 谈群,李志猛,彭黎,等. 天基信息系统任务流度量方法及其应用[J]. 计算机工程与应用 2010,11(32):212-215.
TAN Qun, LI Zhimeng, PENG Li, et al. Space-based information system task-flow measurement method and its application[J]. Computer Engineering and Application, 2010, 11(32):212-215. (in Chinese)

[2] 谈群,彭黎,李志猛,等. 一种航天侦察任务-资源匹配的负载均衡方法[J]. 国防科技大学学报,2011,4(2):95-99.

TAN Qun, PENG Li, LI Zhimeng, et al. A load balancing method for matching reconnaissance tasks and satellite resources [J]. Journal of National University of Defense Technology, 2011, 4(2):95-99. (in Chinese)

[3] 王济川,郭志刚. Logistic 回归模型——方法与应用[M]. 北京:高等教育出版社,2001.
WANG Jichun, GUO Zhigang. Logistic regression model——method and application[M]. Beijing: High Education Press, 2001. (in Chinese)

[4] Jennie P, Simon F. Evaluating the predictive performance of habitat models developed using Logistic regression [J]. Ecological Modelling, 2000,133:225-245.

[5] 贺仁杰. 成像侦察卫星调度问题研究[D]. 长沙:国防科学技术大学,2004.
HE RenJie. Imaging reconnaissance satellite scheduling problem research [D]. Changsha: National University of Defense Technology, 2004. (in Chinese)

[6] 祝江汉,黄维. 基于协商对策的卫星资源冲突消解方法研究[J]. 计算机仿真,2008,25(10):86-89.
ZHU Jianghan, HUANG Wei. Conflict resolution in the satellite mission planning based on bargaining game [J]. Computer Simulation,2008,25(10):86-89. (in Chinese)

[7] 何晓群. 实用回归分析[M]. 北京:高等教育出版社,2008.
HE Xiaqun. Practical regression analysis[M]. Beijing: High Education Press, 2008. (in Chinese)

[8] 刘文卿. 实验设计[M]. 北京:清华大学出版社,2005.
LIU Wenqing. Design of experiment [M]. Beijing: Tsinghua University Press,2005. (in Chinese)

[9] Hosmer D W, Lemeshow S. Applied logistic regression[M]. New York; John Wiley & Sons, 1989.

[10] Scott M. Applied logistic regression analysis [M]. London: Sage Publication, 2002.

[11] Chatterjee S, Hadi A S. Regression analysis by example 4th[M]. New York; John Wiley & Sons, 2006.