

空间信息栅格体系与服务聚合技术*

唐宇¹, 何凯涛^{1,2}, 陈萃¹, 景宁¹

(1. 国防科技大学 电子科学与工程学院 湖南长沙 410073; 2. 中国地质调查局 北京 100035)

摘要 空间信息应用中存在的数据海量、分布异构、处理复杂、计算密集等问题制约了空间数据和信息的广泛共享和应用。空间信息栅格(SIG)以 Grid 和 Web Services 技术为核心,共享并汇集地理上分布的海量空间信息资源,实现动态整合与集成。在构建开放性 SIG 七层应用体系的基础上,重点研究 SIG 服务聚合技术,基于 Petri 网和图论提出了一种新的服务聚合流程模型(服务/资源网),同时构建了可扩展的服务动态选择模型(SISM),实现服务聚合过程中的动态服务选择。结合国家地质调查业务应用,通过地下水资源评价应用实例对 SIG 的实用性和有效性进行了验证。

关键词 空间信息栅格 服务聚合 动态服务选择 国家地质调查

中图分类号:TP393 文献标识码:A

The Architecture and Service Composition of Spatial Information Grid

TANG Yu¹, HE Kai-tao^{1,2}, CHEN Luo¹, JING Ning¹

(1. College of Electronic Science and Technology, National Univ. of Defense Technology, Changsha 410073, China;

2. China Geological Survey, Beijing 100035, China)

Abstract: Some problems in the application of spatial information, such as distributed enormous data, heterogeneous format, complexity of processing and intensive computing have been obstacles to realize wide sharing and integration of spatial information. By taking grid technology and web services as its technical core, spatial information grid (SIG) shares and aggregates the distributed enormous spatial information resources to implement dynamic integration and collaboration. Based on building open seven-layer application architecture of SIG, we put, in the research, emphasis on SIG service composition. A novel service composition model named service/resource net (SRN) is proposed based on Petri net and graph theory. Moreover, an innovative service selection model (SISM) is presented to implement the dynamic service selection in the process of SIG service composition. In accordance with the application requirements of the national geological survey, an application example of ground water evaluation is given and the validity and usability of SIG are verified.

Key words: spatial information grid; service composition; dynamic service selection; national geological survey

空间信息是一种重要的基础性资源,它具有三个区别于其他类型信息的特征:空间特征、专题特征、时间特征^[1]。与空间信息的特征相对应,当前的空间信息应用存在数据海量、分布异构、处理复杂等问题,空间信息无法得到广泛的应用。传统的空间信息技术和系统无法解决现有的应用问题,需要新的技术和方法来管理、整合、处理海量分布的各种空间信息,建立协同的空间信息应用环境,实现一体化的空间信息应用与服务。

作为解决分布式复杂异构问题的新一代技术,网络技术和 Web Services 技术支持资源的发布、发现、调度、访问、监控和应用环境构造,从而实现地理上广泛分布的高性能计算资源、海量数据和信息、数据获取和分析处理系统、应用系统、信息服务以及组织、人员等各种资源的聚合^[2,3]。另一方面,OpenGIS 规范作为空间信息互操作标准,为解决空间数据、信息系统共享和互操作问题提供了一套实用的技术方

* 收稿日期 2004-09-06

基金项目 国家 863 基金资助项目(2002AA131010, 2002AA134010, 2003AA135110)

作者简介 唐宇(1977—),男,博士生。

案^[4,5]。基于此,我们以网格技术、Web Services 技术和 OpenGIS 规范为技术核心,提出空间信息栅格(Spatial Information Grid, SIG)这一概念和技术体系^[6],用于解决当前空间信息应用中的各种问题,满足多层次的空间信息应用需求。

定义 1 空间信息栅格(SIG)是一种汇集和共享地理上分布的海量空间信息资源,并对其进行一体化组织与处理,从而具有按需服务能力的空间信息基础设施。

SIG 是一个复杂的系统,涉及很多关键技术和基本理论的研究,其中,体系结构是 SIG 研究的基础。此外, SIG 以“服务”为中心,应用和交互都基于“SIG 服务”进行(SIG 服务是以独立于平台的方式提供的,可通过接口访问一组空间信息操作, SIG 服务实际是一类特殊的 Web 服务),服务聚合是实现 SIG 应用的关键。结合目前的研究基础,本文重点对 SIG 体系结构、SIG 服务聚合中的流程建模与服务动态选择技术进行研究。

1 SIG 体系结构

SIG 是新一代的空间信息基础设施,它从技术层面上实现分布式空间信息资源的共享和整合。SIG 通过网络将空间信息获取系统、空间数据资源、空间分析处理服务与计算资源(高性能计算机、服务器、PC 等)等连接起来,为空间信息用户提供基于全网空间信息资源的高效服务和综合应用。在 SIG 的研究中,体系结构的研究与设计是最为重要的组成部分之一,它描述和定义了 SIG 系统的构成和层次关系,为构建 SIG 应用提供理论和技术支持,能够保证 SIG 的顶层设计质量。

结合空间信息应用流程(用户请求→空间信息资源发现、聚合→处理任务→返回结果),设计 SIG 应用体系为 7 层结构,称为 SIGOAA(SIG Open Application Architecture)。在此体系中,上层可以调用下层的功能和服务,形成了从空间信息获取、传输到共享、应用的完整体系(如图 1)。

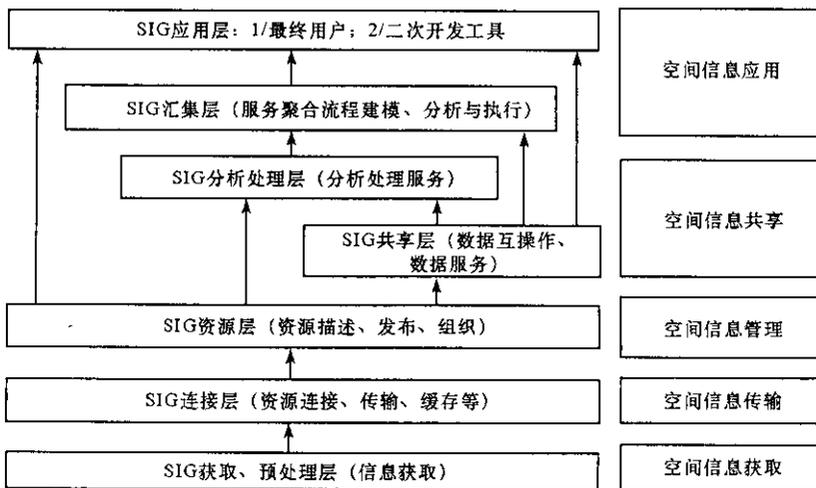


图 1 SIG 应用体系结构

Fig.1 The application architecture of SIG

基于 SIG 应用体系构建 SIG 系统体系(如图 2)。SIG 系统体系由四大部分组成:SIG 资源、资源接入、SIG 核心平台和空间信息应用。其中,资源接入主要是面向不同资源类型,提供相应的 SIG 资源封装器,用于对 SIG 资源细节进行一体化的描述和封装。SIG 资源通过 SIG 资源接入进行封装后以“服务”的方式提供给资源请求者使用。SIG 核心平台主要由 SIG 资源信息服务、SIG 资源管理、SIG 服务聚合与协同引擎、服务调度、安全与监控等功能软件组成,通过一体化的 SIG 资源管理,实现对接入到 SIG 系统的空间信息资源的有效共享和集成。在此基础上,各种空间信息应用(如空间分析处理、数据获取、空间统计等)得以实现。

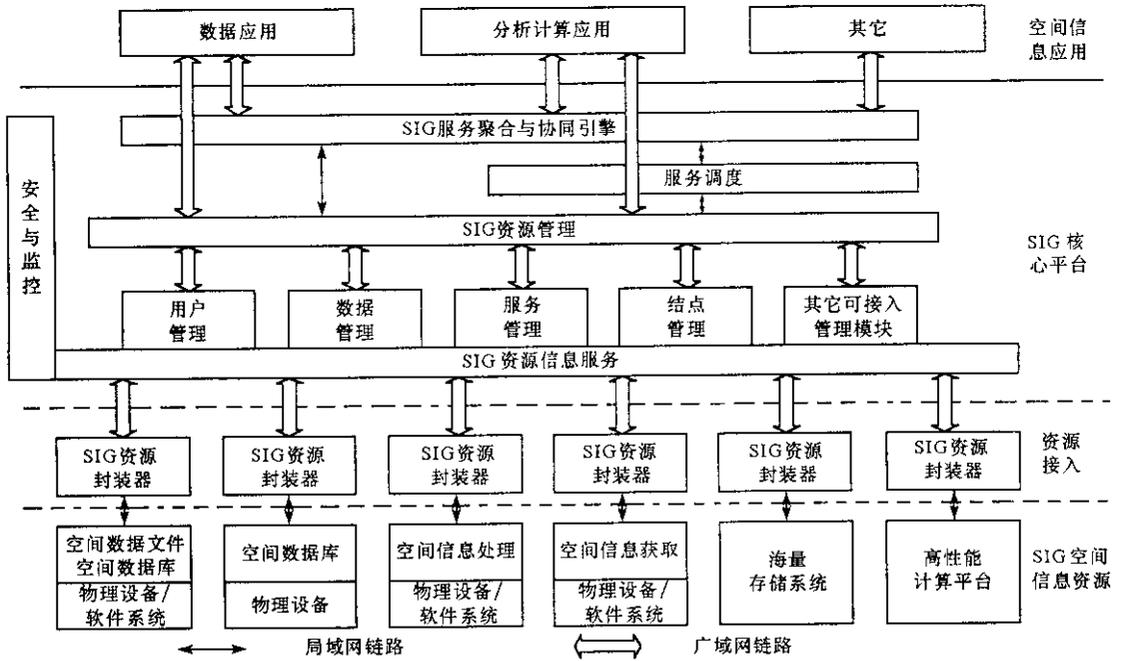


图2 SIG 系统体系结构

Fig.2 The system architecture of SIG

2 SIG 服务聚合流程建模与服务动态选择

在实际的空间信息应用中,单一的 SIG 服务通常无法满足应用需求,需要集成多个服务,形成服务聚合流程来完成任务。SIG 服务聚合是指对现有的一组 SIG 服务按照一定的业务逻辑进行组合,构建执行流程,从而形成新的复杂的复合服务。SIG 服务聚合涉及服务发现、流程建模、流程执行等复杂的过程和技术。本文重点研究 SIG 服务聚合流程建模与聚合流程执行中的服务动态选择技术。

2.1 SIG 服务聚合流程建模

服务聚合流程建模是 SIG 服务聚合的基础。服务聚合流程模型必须能够控制各种复杂情况,这就需要一种形式化的方法和模型来精确地定义和分析服务聚合流程,同时这种方法和模型还要简单易用。Petri 网是一种为描述与分析复杂系统而设计的有效模型工具,它在描述并发、冲突、同步等重要行为现象上具有明显的优势^[7]。因此,引入 Petri 网作为 SIG 服务聚合流程建模的基础。

根据 SIG 服务聚合流程的特点和要求,流程建模过程中需要描述 SIG 服务分类信息,同时还需要支持动态性和条件控制,而简单的 Petri 网无法满足这些应用需求。基于此,对基本的 Petri 网进行扩展,引入分类、时间、条件控制等元素,建立一种适合描述 SIG 服务聚合流程的模型。我们提出的这种扩展模型称为服务/资源网(Service/Resource Net, SRN),其形式化定义如下:

定义2 服务/资源网是一个多元组,即 $SRN = (P, T, F, K, O, I, CLR, CLS, AC, CN, TM, W, M)$ 其中: P 是有限位置集, T 是有限变迁集, F 是表示流关系的集合; K 是位置容量函数, O 是源点, I 是汇点, CLR 是资源分类函数, CLS 是服务分类函数, AC 是流关系弧上的标记函数, CN 是条件约束函数, TM 是 T 上的时间函数, W 是权函数, M 是标识函数。

SRN 是由位置、变迁和令牌、有向弧线所组成的有向图。 SRN 的执行将根据初始标识通过触发变迁得以完成,具体的变迁触发规则和变迁结构参见文献[8]。此外,使用 SRN 进行 SIG 服务聚合流程建模还可以对流程进行分析和性能评价。

2.2 SIG 服务动态选择

SIG 服务聚合流程执行过程由多个服务结点组成(服务结点与 SRN 模型中的“位置”相对应),每个服务结点需要分别完成相应的功能。在实际应用中,对于某种特定功能,可能存在多个 SIG 服务集合与

之相对应;另一方面,在相应的服务集合中,能够提供相应功能的服务通常也不是惟一的,这些服务往往在服务属性和质量等方面存在很大差异,需要根据实际应用情况和一系列规则来进行服务的选择。结合应用实际,服务动态选择问题的解决分为两个步骤:

(1)根据服务聚合流程逻辑,为每一个服务结点选定一个 SIG 服务集合,满足功能需求;

(2)制定一组规则并基于此在各个已选定的服务集合中分别选择一个服务作为服务结点的具体实例供服务聚合流程调用。

与如上步骤相对应,我们首先提出了服务族(Service Family, SF)的概念,用于实现 SIG 服务集合的选择,服务族定义如下:

定义 3 服务族(SF)是指由不同服务提供者提供的、具有相同调用接口、能够实现相同(或相似)功能的一组 SIG 服务, SF 表示为五元组,即 $SF = (N, FD, T, I, O)$, 其中, N 表示 SF 的名称, FD 表示 SF 的功能描述, T 表示 SF 的类型, I 表示输入, O 表示输出。

SF 是 SIG 服务的集合, SF 中的元素可以动态增加和删减,即按照功能描述将 SIG 服务进行归类。由 SF 定义,结合 SIG 服务本体(SIGonto)^[9],计算待选服务族与聚合流程角色功能需求的语义相似度(即服务功能满足聚合流程结点功能需求的程度),从而实现服务结点与服务族的匹配。语义相似度的计算过程如下:对本体库中的概念进行泛化,形成功能需求向量和向量,得出每个服务族的中心向量(该族中所有服务向量简单的算术平均),计算功能需求向量和每个服务族中心向量的余弦相似度,向量夹角越小(即夹角余弦值越大),表明其相似度越高(即服务族的功能越满足功能需求)。选择相似度最高的 SF 作为与服务聚合流程结点匹配的服务族。

对应选定的 SF,需要选择该 SF 中的一个服务实例作为服务结点的调用实例。基于此,引入服务质量要素^[10],建立服务实例动态选择模型(Service Instance Selection Model, SISM),通过定义一组选择规则(包括服务级别定义、服务可靠性、应用概率等)来实现服务实例的选择。

定义 4 服务实例动态选择模型是一个六元组, $SISM = (D, T, C, IP, R, UE)$ 其中:

- Degree(D)表示服务的等级, $D(s)$ 由服务提供者的身份和相关属性决定, $D(s) \rightarrow \{1, 2, \dots, n\}$;
- Time(T)表示服务执行时间(从发出服务调用请求到获得服务处理结果的时间),对于一个 SIG

服务 s 而言, $T(s)$ 的计算将基于先前的服务执行情况与历史数据进行,即 $T(s) = \frac{\sum_{i=1}^n T_i(s)}{n}$, 其中, $T_i(s)$ 表示该服务某一次的执行时间, n 表示该服务的执行次数;

- Cos(C)表示服务请求者调用、执行服务所需支付的费用,服务价格由服务提供者给出;
- Invocation Probability(IP)表示服务的调用概率,设有一个 SIG 服务 s ,其调用概率的计算根据基于先前服务调用的统计数据,即 $IP(s) = N(s)/k$, 其中, $N(s)$ 是服务 s 被调用的次数, k 表示 s 所在的服务被调用的总次数;

• Reliability(R)表示服务的可靠性,即服务在规定时间内调用成功并执行的概率,可靠性与硬件和软件的设置以及网络连接的状况都有关系;其计算依据也是先前服务调用的统计数据, $R(s) = 1 - \frac{M(s)}{N(s)}$, 其中 $M(s)$ 是服务 s 被调用失败的次数, $N(s)$ 表示 s 被调用的总次数;

• User Evaluation(UE)表示用户对服务的评价,其计算依据是调用过服务的用户评价反馈统计数据,即 $UE(s) = \frac{\sum_{i=1}^n UE_i(s)}{n}$, $UE_i(s)$ 表示第 i 个用户的服务评价, $UE(s) \rightarrow \{0, 1, 2, \dots, 100\}$ 。

SISM 模型为服务实例的动态选择提供了计算元素,基于 SISM 进行服务实例选择实际上可以抽象为多属性决策分析问题(Multiple Attribute Decision Making, MADM)^[11]。对于 SIG 服务动态选择过程而言,待选的服务实例构成了决策方案, SISM 的六个元素(D, T, C, IP, R, UE)构成了一组目标准则,计算出各个服务实例在(D, T, C, IP, R, UE)下的属性值;在此基础上,结合服务选择准则,就可以选择出最终的服务实例。SIG 服务动态选择算法设计如下:

step 1: 构建 SIG 服务动态选择的多属性决策矩阵^[11];

step 2: 基于多属性决策矩阵,分析各选择准则之间的关系,以此作为服务排序选优的基础(此处应用熵值法来计算所有方案对某目标准则的贡献总量);

step 3: 结合选择准则对各待选服务进行排序,完成最优服务的选取(此处采用线性分配法进行排序选优,根据目标准则关系(权重)和某目标准则下各方案排序信息构造排序优势度矩阵,然后建立整数规划模型求取排序优势度大的方案^[12]);

3 SIG在地质调查中的应用实例

国家地质调查工作主要是开展公益性、基础性地质调查和战略性矿产资源勘查,为国家经济安全和可持续发展提供保障,为社会提供公益性信息服务。SIG在地质调查应用中的作用可归纳为:以服务的方式实现地质调查应用中的数据共享和软件、应用系统共享,并且基于服务聚合技术,集成各个SIG服务,形成强大的空间信息查询、共享、分析处理能力,满足不同层次的应用需求。

作为地质调查业务应用中的一个重要组成部分,地下水资源评价包括资源量、水环境质量、地下水功能等方面的评价,是一个庞大而复杂的数据处理/评价过程。选择地下水资源评价作为应用实例来分析SIG在地质调查中的应用。在SIG框架下,地下水资源评价过程作为一个服务聚合流程进行应用,SIG数据服务、分析处理服务、计算服务和资源聚合技术在这个过程中分别提供相应的技术支撑,基于SIG的地下水资源评价应用流程如图3所示。

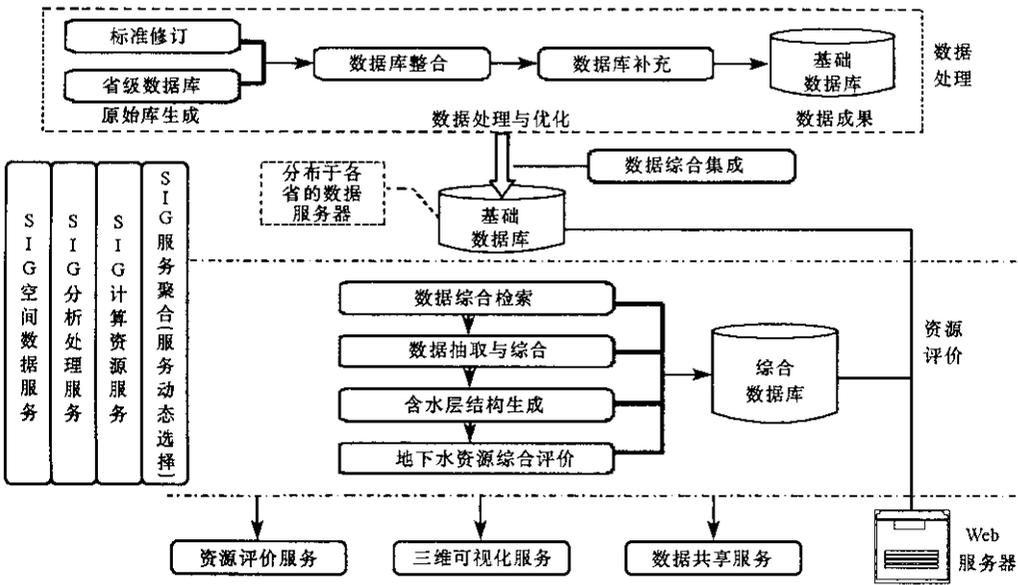


图3 基于SIG的地下水资源评价应用

Fig.3 Ground water evaluating application based on SIG

选定华北平原为应用示范区,以地下水质量标准(GB/T 14848-93)为依据,在三维地下水模型的基础上采用综合指数法提供地下水水质综合评价服务,评价数据源位于不同的数据结点(河北、天津、北京、河南,分别提供相应的数据服务),各种分析处理功能也以服务的方式提供应用。图4表示此实例的应用流程(a)在SIG注册中心(UDDI)查找服务(b)集成分布的数据服务,生成三维地下水模型(c)基于三维模型进行地下水均衡计算(d)综合不同的分析处理服务,完成水质评价,返回分析结果。

由此实例可以看出,SIG能够整合海量分布的各种空间信息资源,建立协同的空间信息应用环境,实现广泛的空间信息共享和应用服务聚合,为实际业务应用提供良好的技术支撑。

4 结束语

SIG为各类空间信息用户提供了一体化的空间信息获取、组织、处理、服务的基本框架和应用环境。

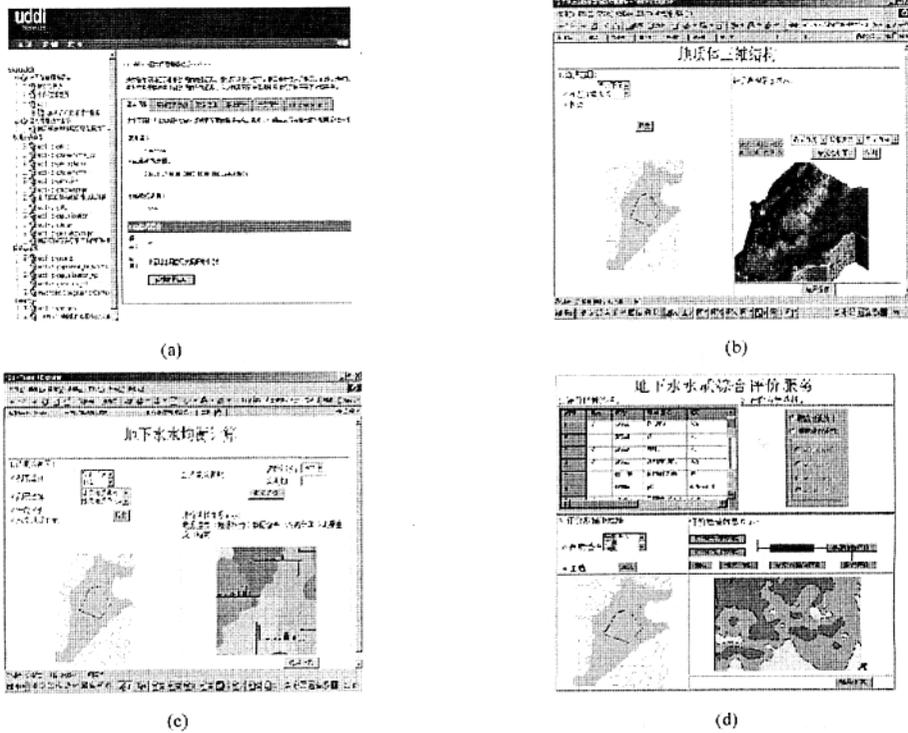


图4 SIG应用实例

Fig.4 An application example of SIG

SIG 是一个探索性、创新性很强的研究课题,涉及很多前沿技术和基本理论的研究。本文提出并建立了开放性的 SIG 体系结构,构建了一种新的 SIG 服务聚合流程模型(SRN),提出了可扩展的 SIG 服务动态选择模型(SISM),并引入多属性决策方法计算服务质量的评价值,实现服务动态选择,同时,面向国家地质调查应用,开发了包括各种典型空间信息服务的应用演示系统。

参考文献:

- [1] 龚健雅. 地理信息系统基础[M]. 北京: 科学出版社, 2001.
- [2] Foster I, Kesselman C, Nick J M, et al. The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration[R]. http://www.gridforum.org/ogsi-wg/drafts/ogsa_draft2.9_2002-06-22.pdf 2002.
- [3] World Wide Web Consortium. Web Services Architecture, Draft 14[R]. <http://www.w3.org/TR/ws-arch/> 2002.
- [4] OpenGIS Consortium. The OpenGIS Abstract Specification Topic 0: Abstract Specification Overview, Version 4.1[R]. <http://www.opengis.org/docs/99-100r1.pdf>, 1999.
- [5] 钟志农, 李军, 景宁, 等. 数字图书馆中地理信息系统的设计与实现[J]. 国防科技大学学报, 2002, 24(6): 87-90.
- [6] 唐宇, 陈萃, 何凯涛, 等. 空间信息栅格 SIG 框架体系与关键技术研究[J]. 遥感学报, 2004, 8(5): 425-433.
- [7] Murata T. Petri Nets: Properties, Analysis and Application[J]. Proceedings of the IEEE, 1989, 77(4): 541-579.
- [8] Tang Y, Chen L, He K T, et al. SRN: An Extended Petri-net-based Workflow Model for Web Services Composition[A]. In Proceedings of the 2nd International Conference on Web Services[C], San Diego, USA, IEEE Press, 2004: 591-599.
- [9] Tang Y, Chen L, He K T, et al. SIG: A Service-oriented Application Grid for Spatial Information Sharing and Integrating[A]. In Proceedings of the 2004 International Conference on Parallel and Distributed Processing Techniques and Applications[C], Nevada, USA, CSREA Press, 2004: 915-922.
- [10] Zeng L, Benattallah B, Dumas M, et al. Quality Driven Web Services Composition[A]. In Proceedings of the 12th International Conference on World Wide Web[C], Budapest, Hungary, ACM Press, 2003.
- [11] 岳超源. 决策理论与方法[M]. 北京: 科学出版社, 2003.
- [12] 束金龙, 闻人凯. 线性规划理论与模型应用[M]. 北京: 科学出版社, 2003.

