

基于网格技术的空间知识发现与数据挖掘研究

葛小三¹

(1 武汉大学空间信息与数字工程研究中心, 武汉市珞喻路 129 号, 430079)

摘要: 设计了基于服务网格的空间知识发现与数据挖掘的架构, 并对其关键技术进行了探讨性研究。

关键词: 知识发现; 数据挖掘; 空间知识网格

中图法分类号: P208; TP182

空间知识发现和数据挖掘(SDMKD)是空间信息科学的重要研究内容^[1-3]。它通过综合运用人工智能、计算智能、模式识别、数理统计等技术从大量空间数据中挖掘和发现有价值和隐含的知识。随着以超高速的机器处理能力、复杂的数据存储方法、下一代远程通信能力、新一代操作系统和高级网络服务能力为标志的网格技术^[4]的出现,对空间知识发现和数据挖掘提供了一体化空间知识发现与数据挖掘环境——空间知识网格。空间知识网格是在空间信息网络与语义网格基础框架上建立基于网格技术的空间知识发现和应用服务的有效共享、管理的模型、方法和平台。空间信息网络是网格环境下空间信息获取、处理与应用的基本技术框架,是一体化的空间信息处理平台和基本应用环境^[5,6]。空间语义网格以空间信息网络为空间信息基础设施^[7],在空间信息网络协议的基础上,为用户提供基于语义的一体化空间信息应用服务的网络平台。空间语义网格的目标是把网络上的空间信息服务资源在语义层链接起来,实现基于语义的集成和互操作。对空间信息网络的研究集中在解决不同计算资源的链接上,以便用户可以透明、无缝地使用网络上的所有资源。空间语义网格需要解决同构与异构的信息访问,在此基础上,空间知识网格能进行空间知识发现和高级空间知识应用服务,即空间信息网络完成数据访问途径、语义网格解决空间数据/消息的共享与互操作,知识网格来完成空间知识发现任务^[8-10],这是空间数据与信息利用最高的阶段。

1 空间知识网格中的空间知识流程

空间知识网格中空间知识流是指空间知识根据一定的规则和原则在空间知识流网格中的传递过程。空间知识网格中空间知识流涉及到从空间数据生产、加工、处理及挖掘到空间知识提取以及用户应用知识的整个过程^[11]。

1) 理解和定义问题。理解和定义问题是十分重要的一步,根据需求准确地表达机器所能理解的语义;因为问题的解决是在复杂的网络环境中,在整个知识发现过程中,所有的涉及到的语义共享与互操作,以及由此派生的相关的其他对知识发现过程的资源发现、选取与约束都要以定义的问题为基础。

2) 选取或建立解决问题的模型。确定知识发现的目标,确定知识发现系统是发现哪种类型的空间知识。

3) 空间数据发现与选取。根据前边两个步骤,要在网格环境中进行相关空间数据的动态发现与定位;在空间数据网格相应节点提取数据(包括地形数据、GPS数据、地图数据、影像数据等其他传感器获取的空间数据)。大多数情况下,相关空间数据是从已存贮在数据网格中的各数据节点的数据库或数据仓库中提取的。根据要求从数据节点中提取与空间知识发现相关的数据,形成目标数据源。

4) 空间数据处理与再加工。

5) 确定知识发现算法与工具。根据所要发

现的空间知识类型及动态集成的空间数据,在网格环境中发现空间数据挖掘与知识发现相应的算法与工具。

6) 进行空间数据挖掘与知识发现。在空间知识网格环境下,运用空间数据挖掘工具对动态集成的空间数据进行数据挖掘与知识发现,挖掘出感兴趣的空间知识。这个阶段是空间知识发现中最重要的一环。

7) 知识评价。运用领域相关的知识来解释所发现的空间知识,将其存贮到空间知识库,并对所发现的知识进行评价,看是否能满足要求,如果不能,则改变约束条件,重复步骤 3)~ 步骤 6)。

8) 空间知识的应用。在此基础上进一步与其他非空间知识进行融合,确定模型参数,运用建立的模型解决问题。如果必要,则对结果给出解释,并作出评价。

2 面向服务的空间知识发现架构

网格环境下空间知识发现结构体系是空间数据(资源)发现、计算资源配置、空间数据挖掘与知识发现等各种技术的动态集成的架构体系。空间知识网格研究的重要任务就是建立合理的空间知识发现与数据挖掘的结构体系。

开放的网格体系结构(OGSA)的概念是结合现有的网格标准、面向服务的体系结构以及 Web 等的结果。Web 服务标准为服务的创建、服务的命名以及服务的发现定义了开放式标准的机制。空间知识网格环境下空间知识发现与数据挖掘服务是一种充分利用分布在网格中的空间数据、空间网格服务、计算资源、存储资源等资源来构建分布式空间知识发现与数据挖掘应用(系统)的方法论与技术体系。本文基于 OGSA 建立了网格环境下空间知识发现与数据挖掘的面向服务的体系结构(图 1)。该体系结构包括划分网格环境下空间知识发现与数据挖掘服务的基本组成部分、各部分的功能的定义、各部分间关系的描述以及将它们集成在一起而实现空间知识应用的方法。其中核心服务为:空间数据的访问服务;空间数据处理与动态集成服务;空间数据挖掘与知识发现服务;空间知识应用服务。

空间知识服务涉及到空间知识获取、建模、检索、重用、发布和管理等多个方面。图 2 是对图 1 中空间知识服务层的扩展。在这个扩展服务层中,是将空间知识发现与应用技术领域的各种工具、方法及服务集成为一个整体。这样一个面向

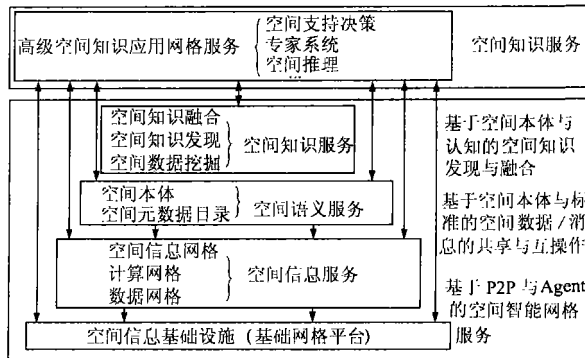


图 1 面向服务的空间知识网格架构

Fig. 1 Service-oriented Architecture of SKG

服务的知识扩展层是为了适应灵活、可扩展且易重用和共享空间知识资源,使空间知识服务具有分布性和适应性。相互链接不同的知识管理任务之间的交互不是固定不变的,可以更新其中的任一组件而不用改变其他的组件,新的技术与工具随时可以采用。空间知识可以随时加入到空间知识库中去。例如,空间数据挖掘服务可以随时加入以便能够自动获取知识和动态更新知识库。

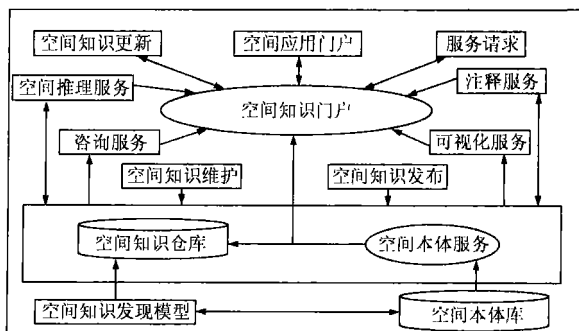


图 2 面向服务的知识扩展层

Fig. 2 Service-oriented Extended Spatial Knowledge Layer of SKG

3 关键技术

空间知识网格中空间知识发现是在一种松耦合网格环境下的空间数据发现、访问、处理、挖掘到空间知识提取以及用户应用知识的全过程,涉及到多个学科的内容和多方面的技术。

1) 智能搜索与资源抽取技术。XML 和 Ontology 在语义 Web 中的使用,对网络搜索技术带来了新的契机。通过本体的定义,搜索程序可以进行基于概念的精确搜索。通过机器学习和人工智能算法,实现基于自然语言理解的空间数据信息抽取的新一代空间智能搜索引擎对空间知识发现将提供有力的保证。用户可以透明地访问操纵

空间数据资源, 而不需要关心数据源的物理位置和结构特性。

2) 面向空间知识发现服务的动态集成技术。

3) 网格环境下空间数据/信息共享与互操作技术。空间数据/信息共享与互操作技术同样是网格环境中将不同来源不同数据组织形式的空间数据/信息进行有效管理和综合应用的难点。由于空间数据的异构性特点、数据内容与来源的差异性、空间数据模型差异性、支撑软件平台的差异性, 如何根据特定的网格服务使这些集成数据能够在网格环境下的不同系统中相互可操作以及异构分布的空间数据共享, 对于空间数据挖掘与知识发现十分重要。

4) 新的挖掘与知识发现算法与工具。

5) 空间知识与数据可视化技术。网格环境下空间数据挖掘与知识发现可视化技术包括空间数据可视化技术、挖掘与发现结果可视化技术、挖掘过程可视化技术等一系列的数据挖掘与知识发现可视化技术。空间知识与数据可视化技术可以提高用户对空间知识与数据的理解, 能极大地改善系统挖掘速度和深度, 从而增强提取新的有用的空间知识的能力。

6) 信息安全技术。在网格环境下, 安全是指允许合法用户合法使用网格中的资源。空间数据挖掘与知识发现涉及到网格中空间数据访问与传输的安全技术, 同时也涉及到空间数据挖掘与知识发现过程中隐私的保护问题。

参 考 文 献

- [1] Foster I, Kesselman C. The Grid 2: Blueprint for a New Computing Infrastructure [M]. San Francisco: Morgan Kaufmann, 2004: 20-28
- [2] 邱凯昌, 李德仁, 李德毅. 空间数据挖掘和知识发现的框架[J]. 武汉测绘科技大学学报, 1997, 22(1): 328-332
- [3] 李德仁, 王树良, 史文中, 等. 论空间数据挖掘和知识发现[J]. 武汉大学学报·信息科学版, 2002, 27(3): 221-233
- [4] Joseph J, Fellenstein C. Grid Computing [M]. USA: IBM Software Press, 2004: 1-6
- [5] Luo Yingwei, Wang Xiaolin, Xu Zhuoqun. Spatial Information Grid: an Agent Framework [C]. The Second International Workshop, GCC 2003, Shanghai, 2003
- [6] Ren Yan, Fang Cheng, Chen Honghui, et al. Research on the Application of Multi-agent Technology to Spatial Information Grid [C]. GCC 2003, Shanghai, 2003
- [7] 李德仁, 崔魏. 空间信息语义网格[J]. 武汉大学学报·信息科学版, 2004, 29(10): 847-851
- [8] Congiusta A, Mastroianni C, Pugliese, et al. Enabling Knowledge Discovery Services on Grids [C]. AxGrids 2004, LNCS, 2004
- [9] Buetti G, Congiusta A, Talia D. Developing Distributed Data Mining Applications in the KNOWLEDGE GRID Framework [C]. VECPAR 2004, LNCS, 2005
- [10] Berman F. From Tera Grid to Knowledge Grid [J]. Communications of the ACM, 2001, 44(11): 27-28
- [11] Di Liping, Distributed Geospatial Information Services Architectures, Standards, and Research Issues [OL]. <http://www.isprs.org/istanbul2004/comm2/papers/121.pdf>, 2006

[1] Foster I, Kesselman C. The Grid 2: Blueprint for a New Computing Infrastructure [M]. San Francisco: Morgan Kaufmann, 2004: 20-28

作者简介: 葛小三, 博士生, 现从事 GIS、空间数据挖掘与知识发现研究。

E-mail: gexiaosan@163.com

On Spatial Knowledge Discovery and Data Mining in Grid Environment

GE Xiaosan¹

(1 Research Center of Spatial Information and Digital Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

Abstract: The issue of spatial data mining and knowledge in the grid environment are introduced, the produce of spatial knowledge discovery, sharing and fusing, extraction and application is described, an architecture for integrate spatial knowledge application in grid environment is designed, the key technologies of spatial knowledge grid also are discussed.

Key words: knowledge discovery; data mining; spatial knowledge grid

About the author: GE Xiaosan, Ph D candidate, majors in GIS, SDMKD.

E mail: gexiaosan@163.com