

# 通用知识网格下以用户为中心的数据挖掘本体研究<sup>\*</sup>

李玉华 卢正鼎

(华中科技大学计算机科学与技术学院计算机科学与技术系 武汉 430074)

**摘要** 分布异构的海量数据挖掘是数据挖掘领域亟待解决的课题,通用知识网格(UKB)架构模型用于在网格环境下创建大规模的分布式知识发现和知识集成系统。本体服务器是整个架构的核心模块,负责本体的管理和查询。数据挖掘本体服务是本体服务器提供的主要服务。本文主要介绍通用知识网格下以用户为中心的数据挖掘本体的设计和 OWL 实现。数据挖掘本体可满足各种不同领域、不同层次用户的知识发现服务,使系统具有开放性、可扩展性和高用户可用性。还介绍了一个反洗钱领域数据挖掘解决方案实例。

**关键词** 以用户为中心,数据挖掘,本体,通用知识网格

## The User-centered Data Mining Ontology Development on Universal Knowledge Grid

LI Yu-Hua LU Zheng-Ding

(College of Computer Science, Huazhong University of Sci. & Tech., Wuhan 430074)

**Abstract** There are many needs for solving large-scale distributed knowledge discovery problem. The Universal Knowledge Grid (UKG) is an ontology-based grid architecture model to build large-scale distributed knowledge discovery system on the grid. The data mining ontology services are the main service offered by UKG. This paper mainly introduces the development and OWL implementation of the user-centered data mining service ontology on Universal Knowledge Grid. The data mining ontology services can meet the user requirements of knowledge discovery in different domains and different hierarchies and make the system exoteric, extensible and high usable. A data mining solution for money laundering is introduced.

**Keywords** User-centered, Data mining, Ontology, UKG

### 1 引言

随着计算机、互联网等技术的迅速发展,各种信息数据呈爆炸式增长。当今数据库的容量已经达到 TB 级和 PB 级,人们淹没在数据的海洋里,而用于决策支持的有用的知识却无法获得。那么怎么得到这些“知识”呢?计算机科学对这个问题给出的最新回答就是:数据挖掘,在“数据矿山”中找到蕴藏的“知识金块”,帮助企业减少不必要投资的同时提高资金回报。

数据挖掘是一个利用各种分析工具在海量数据中发现模型和数据间关系的过程,这些模型和关系可以用来做出预测。

在许多科学和商业领域,有很多 TB 级和 PB 级的数据集有待分析;而且在某些情况下数据集是分布在一个组织的在不同地点或不同组织的多个数据源集成且必须被集团内所有用户共享。

要解决这种大规模的分布式知识发现和知识集成系统问题,需要一种便于资源集成、数据挖掘服务、知识共享和知识集成,并具有较高的安全性的架构模型。

网格(grid)<sup>[1]</sup>是构筑在互联网上的一组新兴技术,实现计算资源、存储资源、信息资源、知识资源等的全面共享,消除信息孤岛和资源孤岛。网格试图实现互联网上所有资源的全面连通。

本体<sup>[2]</sup>最早是一个哲学上的概念。从哲学的范畴来说,

本体是客观存在的一个系统的解释或说明,关心的是客观现实的抽象本质。Studer 等认为本体是共享概念模型的明确的形式化规范说明。本体的目标是捕获相关领域的知识,提供对该领域知识的共同理解,确定该领域内共同认可的词汇,并从不同层次的形式化模式上给出这些词汇(术语)和词汇间相互关系的明确定义。

通用知识网格<sup>[3]</sup>(Universal Knowledge Grid, UKG)是一种基于本体的网格架构模型,可在网格环境下提供高性能的知识发现和知识集成服务,为分布异构海量数据的知识发现和知识集成系统提供一个新的应用解决方案。本体服务器是整个架构的核心模块,负责本体的管理和查询。数据挖掘本体服务是本体服务器提供的主要服务。

以用户为中心的设计被认为特别强调使产品具有可用性的过程。产品的可用性<sup>[4]</sup>在 ISO 9241-II 中给出的定义是“产品能够被特定使用者为特定目标在特定使用环境下方便、有效、满意地使用的程度。”其目的是指导产品设计以便使用者可以凭直觉使用它们。我们开发产品或者软件,先要搞清楚用户的水平,在此基础上才能开发出高可用性的产品。

在传统的设计中,KDD 解决方案经常特别重视算法、架构、软件、硬件和系统工程挑战,而没有首先探讨最终用户将如何使用新的 KDD 技术。在这种以系统为中心的设计中,系统实现了许多没用的特征而延长了系统周期,使系统难于操作和使用。

<sup>\*</sup>国家自然科学基金(60403027)资助项目、国家“十五”重大科技攻关(2001BA102A06-11)资助项目。李玉华 博士研究生,副教授,主要研究方向为知识管理、数据挖掘、语义 Web 和本体论。卢正鼎 教授,博士生导师,主要研究方向为 CIMS、分布式异构系统集成。

本文主要介绍通用知识网格下以用户为中心的数据挖掘本体的设计和实现,可满足各种不同领域、不同层次用户的知识发现服务,可缩短开发周期、降低消耗并具有较好的可用性,使系统具有开放性、可扩展性和高用户可用性。

## 2 通用知识网格的架构模型

首先介绍通用知识网格的架构模型<sup>[3]</sup>,如图 1 所示,UKB 是定义在基本的计算网格工具集和网格服务基础上的。下面分别介绍各部分的主要功能。

### 2.1 智能知识浏览器

智能知识浏览器提供一组图形化的工具,用户使用可视化的工具或自然语言创造自己的知识需求。智能知识浏览器可显示机器可理解的语义并解释显示的内容。用户可按语义操作资源,具有自动推荐服务、高级导航、推理能力。

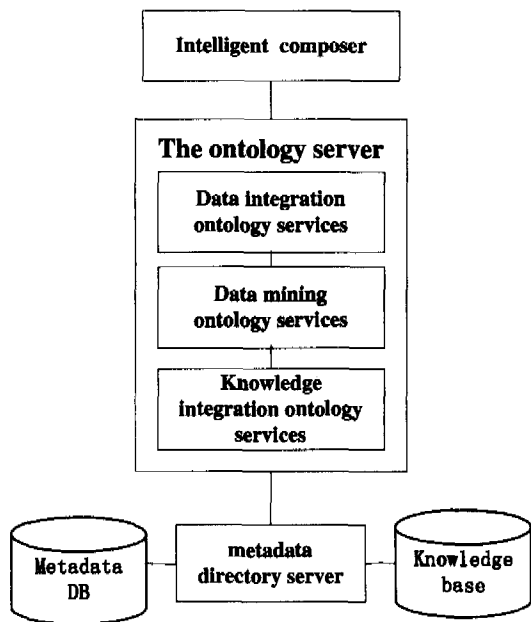


图 1 通用知识网格的架构

### 2.2 本体服务器

本体服务器是整个架构的核心模块,负责本体的管理和查询。本体服务器提供下面的服务:数据集成本体服务、数据挖掘本体服务、知识集成本体服务、安全策略本体服务、防火墙本体服务。下面分别加以介绍。

**2.2.1 数据集成本体服务** 数据集成本体服务描述 Web 文档的语义,在半结构化数据和结构化数据库之间架起一座桥梁,帮助进行数据清洗和数据准备,进行异种数据源集成。

**2.2.2 数据挖掘本体服务** 数据挖掘本体服务可给用户 提供数据挖掘应用需求、应用解决方案和算法资源语义模型,为用户提供高可用性的数据挖掘服务。

**2.2.3 知识集成本体服务** 知识集成本体服务包括以下功能:

- ①维护公共有效本体集;
- ②负责查询条目之间的关系;
- ③方便不同知识库之间的通讯和知识交换;
- ④不同本体之间的翻译和映射;
- ⑤对数据挖掘本体挖掘的模式进行评分,并把有效的知识补充到知识库;
- ⑥提供按需服务,以支持问题解决和决策支持;

⑦集成不同水平的知识源,以支持问题的分析和解决。

### 2.3 元数据目录服务器

元数据目录服务器维护 UKB 中描述的所有数据、工具和知识的元数据。元数据信息用 XML 文档表示,存储在元数据库和知识库中。

### 2.4 元数据库

元数据库存储着各种数据源、数据集成和数据挖掘工具、算法的元数据。

### 2.5 知识库

知识库存储应用领域收集的各种知识及数据挖掘的结果,如学习的模型和发现的模式等。

## 3 数据挖掘本体的构造

要创建 UKB 的数据挖掘本体服务,需深入研究数据挖掘技术及数据挖掘应用系统的研究现状和发展趋势,确定数据挖掘本体的范围。

### 3.1 数据挖掘本体的设计

人们已经对数据挖掘技术和数据挖掘系统进行了大量研究。Robert Grossman 在 1998 年从技术角度分析提出了四代数据挖掘系统<sup>[5]</sup>的概念;Gregory Piatetsky-Shapiro 在 KDD2000 上从应用角度总结了数据挖掘系统的发展<sup>[6]</sup>,可将其归纳成 3 个阶段:独立的数据挖掘软件,横向的数据挖掘工具集,纵向的数据挖掘解决方案;Lu Hongjun 在 PAKDD'01 进一步提出数据挖掘系统与数据库、数据仓库系统集成的趋势<sup>[7]</sup>;Han Jiawei 在 2001 年提出数据挖掘技术与应用结合开发纵向的数据挖掘系统的发展方向<sup>[8]</sup>;Eric Haseltine 在 KDD2004 上借鉴顾客产品工业的以用户为中心的方法<sup>[9]</sup>,提出开发以用户为核心的 KDD 设计,这种方法可缩短开发周期、降低消耗并具有较好的可用性。

在数据挖掘本体的构造中,充分综合前述人们对数据挖掘的成果及趋势分析,遵循以用户为核心的设计理念,为用户 提供高可用性的数据挖掘服务。数据挖掘本体的结构如图 2 所示。

从技术角度,数据挖掘主要是实现特征提取、概念描述、分类、预测、聚类、链结分析、关联分析、孤立点分析、进化分析、可视化、串比较等功能。一般的数据挖掘算法都是针对结构化的数据,针对复杂的半结构化或非结构化数据等复杂的数据,如复杂数据对象、空间数据、多媒体数据、时间序列数据、文本数据和 Web 信息,需要一些特殊的处理和算法。

数据挖掘从一开始出现就是面向应用的,其应用领域非常广泛。按照 Kdnuggets 公司 2004 年 8 月关于数据挖掘应用领域的网站调查结果<sup>[10]</sup>,我们把数据挖掘应用领域主要分为银行、生物信息、直接销售和筹款、电子商务及 Web、保险、投资和股票、制造业、医药业、零售业、科学数据、安全、电信、旅游、其他等 15 个应用领域,每个领域还可划分为子领域,每个子领域包含多个具体的应用任务。

每个具体的应用任务包含一个或多个数据挖掘算法,以实现相应的应用。每个具体的应用任务的实现,关键是要根据应用特点确定任务所使用的数据挖掘算法。如电子商务及 Web 应用领域又可分为网上商品推荐、个性化网页、自适应网站、客户分析等;客户分析子领域又可包括客户分类、客户行为分析、客户利润分析等业务应用任务。

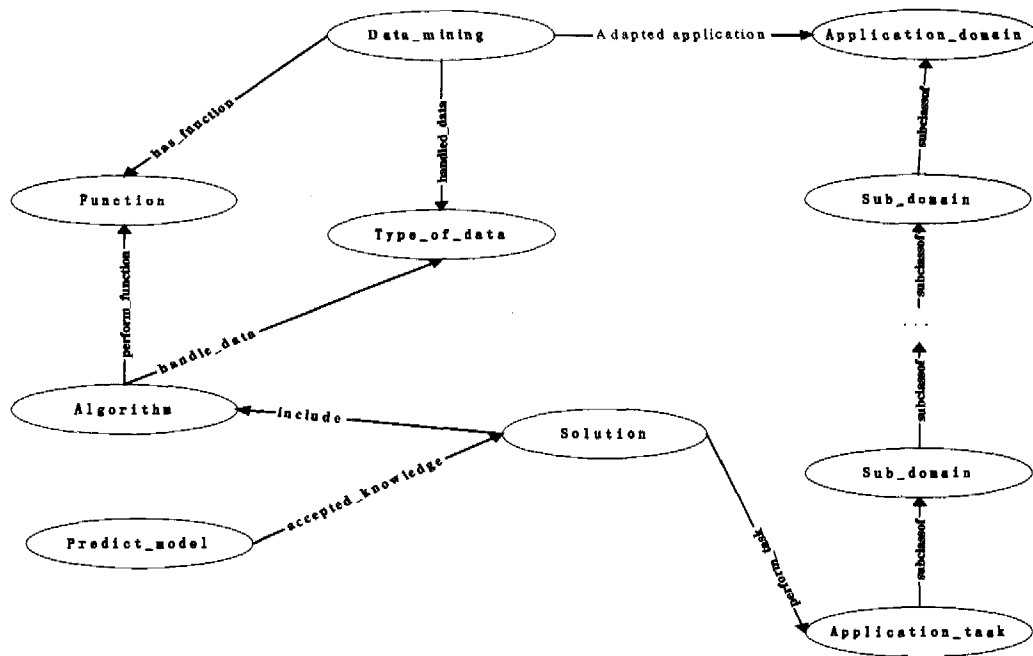


图2 数据挖掘本体结构

实施数据挖掘解决方案,获得知识或预言模型。预言模型的描述和存储采用 PMML<sup>[11]</sup> 标准。PMML 是基于 XML 的预言模型标记语言,用于定义统计和数据挖掘的模型及在 PMML 兼容的应用中共享模型。由 DMG (data mining Group, <http://www.dmg.org>) 组织发布,并已经被 W3C 接受,作为数据挖掘模型定义的标准,当前 PMML 最新版本是 3.0。

### 3.2 本体实现

为了便于 Web 上应用程序使用方便,需要有一个通用的标准语言来表示本体,就像 XML 作为标准的数据交换语言一样。目前正在开发中的语言有:SHOE、OML、XOL、Ri-boweb、RDFS 和 OIL。在本体的编码中采用 W3C 提出的

OWL<sup>[12]</sup> (Web Ontology Language) 语言。

```

<owl:Class rdf:ID="Application_Domain"/>
  <owl:Class rdf:about="#Sub_domain">
    <rdfs:subClassOf rdf:resource="#Application_Domain"/>
    <rdfs:label>Sub domain</rdfs:label>
  </owl:Class>
  .....
  <owl:Class rdf:ID="Application_Task">
    <rdfs:subClassOf>
      <owl:Class rdf:ID="Sub_domain"/>
    </rdfs:subClassOf>
  </owl:Class>

```

图3 描述概念类间的子类关系代码

```

<owl:Class rdf:ID="Data_Mining"/>
  ...
  <owl:FunctionalProperty rdf:ID="handled_data">
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
    <rdfs:domain rdf:resource="#Data_Mining"/>
    <rdfs:range rdf:resource="#Type_of_Data"/>
  </owl:FunctionalProperty>
  <owl:FunctionalProperty rdf:ID="has_function">
    <rdfs:domain rdf:resource="#Data_Mining"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
    <rdfs:range rdf:resource="#Function"/>
  </owl:FunctionalProperty>
  <owl:FunctionalProperty rdf:ID="adapted_application">
    <rdfs:range rdf:resource="#Application_Domain"/>
    <rdfs:domain rdf:resource="#Data_Mining"/>
    <rdf:type rdf:resource="http://www.w3.org/2002/07/owl#ObjectProperty"/>
  </owl:FunctionalProperty>

```

图4 Data mining 类的部分代码描述

OWL(Web Ontology Language)<sup>[14]</sup> 适用于这样的应用:在这些应用中,不仅仅需要提供给用户可读的文档内容,而且希望处理文档内容信息。OWL 能用来清晰地表达词汇表中的词条(term)的含义以及这些词条之间的关系。而这种对词条和它们之间的关系的表达就称作本体。OWL 比 XML、RDF 和 RDFSchem 拥有更多的机制来表达语义,从而超越了 XML、RDF 和 RDFSchem 仅仅能够表达网上机器可读的文档内容的能力。

本体开发工具采用 Stanford 的 protege2000。数据挖掘本体包含数据挖掘(Data mining)、功能(Function)、数据类型(Type\_of\_data)、应用域(Application\_domain)、子域(Sub\_domain)、应用任务(Application\_domain)、解决方案(Application\_task)、算法(Algorithm)、预言模型(Predict\_model)等概念类。概念之间的关系除了应用域和子域、子域和应用任务之间的子类(subclassof)关系外,都是通过概念属性实现的。数据挖掘类用“handled\_data”、“has\_function”、

“adapted\_application” 3 个属性建立了它与数据类型、功能、应用域类之间的关系, 解决方案类用“include”、“perform-task”两个属性建立了它与算法及应用任务类之间的关系, 算法类用“perform\_function”、“handle\_data”两个属性建立了它与功能及数据类型类之间的关系, 预言模型类用“accepted\_knowledge”建立了它与解决方案类之间的关系。

由于空间的限制, 这里只介绍两个概念之间的关系的 OWL 编码表示。图 3 的代码描述了子域是应用域的子类, 应用任务是子域的子类。图 4 的代码描述 Data mining 类的一部分, 描述了数据挖掘类用“handled\_data”、“has\_function”、“adapted\_application”3 个属性实现了它与数据类型、功能、应用域类之间的关系。

### 3.3 洗钱领域数据挖掘解决方案一例

洗钱是国家安全领域的子领域欺诈应用领域的子领域。在国家安全领域<sup>[13]</sup>; 自从 2001 年“911 事件”以后, 对于国家安全的关注受到各国的普遍重视。现在国际化的贩毒、洗钱、恐怖等犯罪活动日益猖獗, 各国都在积极组织人力, 密切监视各种犯罪活动, 以保卫国家安全。国家安全领域可分为交通肇事、性犯罪、盗窃、纵火、贩毒、暴力犯罪、网络犯罪、欺诈等子领域, 每个子领域还可继续划分, 分解成一个个具体业务应用。

洗钱是欺诈应用领域的一个子域。欺诈领域包含洗钱、造伪钞、保险欺诈、腐败和受贿、盗用资产(包括软件、音乐、电影和其他智能产品)子领域。洗钱子领域包括交易网络分析、交易属性相关性分析、不同案件分组、异常资金交易分析、非正常访问序列特征分析、交易趋势分析等业务应用, 具体业务应用与对应的所用数据挖掘算法如表 1 所示。

表 1 洗钱应用解决方案及其所用的数据挖掘算法

业务应用	算法类别	具体算法
交易网络分析	链结分析	Coplink Concept Space <sup>[14]</sup> 、CLIQUE <sup>[15]</sup> 、two-tree PFS <sup>[16]</sup>
交易属性相关性分析	分类	SLIQ <sup>[17]</sup>
不同案件分组	聚类	CLIQUE <sup>[18]</sup>
异常资金交易分析	孤立点分析	TOP-n LOF <sup>[18]</sup>
非正常访问序列特征分析	序列模式分析	multi-objective GA <sup>[19]</sup>
交易趋势分析	预测	Multiple regression <sup>[20]</sup>

数据挖掘本体服务给用户提供了数据挖掘应用需求、应用解决方案和算法资源语义模型。在本体的帮助下, 最终用户通过以下步骤实现一个具体的数据挖掘任务:

1. 利用智能知识浏览器, 用户通过数据集成本体选择数据源, 并输入要解决的具体业务问题, 如交易网络分析;
2. 数据挖掘本体根据语义模型帮助用户找到交易网络分析适用的数据挖掘应用解决方案, 包含 Coplink Concept Space、CLIQUE 和 two-tree PFS 算法;
3. 系统将数据源和应用解决方案信息传给元数据目录服务器, 执行应用解决方案, 获得预言模型, 通过知识集成本体进行评价, 然后测试;
4. 系统将测试后的预言模型存入知识库, 同时通过智能知识浏览器反馈给用户;
5. 用户可将测试的预言模型, 应用到事务处理系统, 辅助支持领导决策;
6. 以后用户还可直接调用知识库的预言模型, 直接用于

或修改后用于事务处理系统, 辅助支持领导决策。

### 3.4 系统设计的特点

1) 用户具有更强的可用性。

用户不需要了解具体的数据挖掘算法, 只需根据自己的业务应用, 调用业务应用解决方案, 实施数据挖掘任务。将挖掘出的预言模型通过知识集成本体的模型评价本体进行评价并测试后, 以 PMML 形式存储在知识库中, 并将预言模型, 应用到已有的业务处理系统, 为决策支持服务。

用户易于编辑预言模型, 因为作为 XML 文档, 可以通过文本编辑器进行编辑。在已有模型的基础上, 根据自己的需求开发适用的新模型, 缩短开发周期, 降低开发费用, 具有更强的可用性。

用户也可根据自己的应用, 利用已有的数据挖掘算法, 方便地定义新的应用解决方案模型。

用户可直接调用知识库中的预言模型, 应用到已有的业务处理系统, 为决策支持服务。

2) 系统的开放性和可扩展性。

PMML 是描述数据挖掘模型的标准, 遵循这种标准, UKG 挖掘出的应用模型能够被其他 PMML 兼容的数据挖掘系统共享, 同时其他数据挖掘系统得出的模型如果存储为 PMML 格式, UKG 也能识别并利用, 实现 UKG 能够与其他数据挖掘系统之间共享模型。PMML 的这些特点, 保证了采用 PMML 这种开放的格式存储, 可以将挖掘模型与数据及挖掘工具分离开。

## 4 相关工作

表 2 示出了网格环境下支持知识发现的不同的系统和方法<sup>[21]</sup>。近来出现的许多面向知识发现的系统, 是为某个特定领域而设计的, 后来被扩展, 以支持较通用的应用创建。这些系统中的一些确实具有集成了访问和处理大型数据集的先连接口, 而且提供了支持典型知识发现过程的专门功能。

Knowledge Grid<sup>[2]</sup> 是一个网格环境下独立于域的知识发现环境, 它提供了专门设计的集成并行和串行的数据挖掘服务, 管理基本的数据集和抽取的知识模型。但系统只有精通数据挖掘算法的专家才能熟练使用, 如果对算法不了解, 难以得出好的知识模型。

UKG 是一种基于本体的网格架构模型, 可在网格环境下提供高性能的知识发现和知识集成服务, 为分布异构海量数据的知识发现和知识集成系统提供了一个新的应用解决方案; 可满足各种不同领域、不同层次用户的知识发现服务, 使系统具有开放性、可扩展性和高用户可用性。

**结论及相关的工作** 本文介绍的以用户为中心的数据挖掘本体服务, 是通用知识网格本体服务器提供的主要服务。它不仅提供丰富的针对各种不同数据挖掘功能和数据类型的数据挖掘算法, 还提供不同应用领域多个应用任务的数据挖掘应用解决方案及挖掘出来的预言模型, 可满足各种不同领域不同层次用户的知识发现和数据挖掘服务, 使系统具有开放性、可扩展性和高用户可用性。

文章还介绍了一个洗钱领域数据挖掘解决方案实例, 包括应用域的层次、数据挖掘应用解决方案和所用算法的映射及用户如何在系统帮助下实现数据挖掘应用。

系统将来的工作就是将更多的应用领域解决方案集成到系统的数据挖掘本体服务, 研究更有效的知识集成服务, 为用户提供高可用性的知识发现和知识集成服务。

(下转第 245 页)

所得到的旋转角度具有鲁棒性强的优点。从提取出的特征点图像中可以看出,本文提出的方法提取的特征点定位非常精确。

**结束语** 在本文提出的基于提出的基于小波多尺度积的图像配准方法,由于提出的边缘图像和特征点具有抗噪性强和精确性高等优点,因此,在特征点对匹配过程中,结合多种匹配准则都可以达到满意的效果和精度。

**参考文献**

- 1 Brown L G. A survey of image registration techniques, ACM Computing Surveys, 1992, 24(4): 325~376
- 2 Zitova B, Flusser J. Image registration methods: a survey, Image and Vision Computing, 2003, 21: 977~1000
- 3 ZHOU Ji, SHI Jiao-ying. A Robust Algorithm for Feature Point Matching, Computers & Graphics, 2002, 26: 429~436
- 4 ATALAY V, YILMAZ M U. A Matching Algorithm Based on Linear Features, Pattern Recognition Letters, 1998, 19: 857~867
- 5 DAI Xiao-long, KHORRAM s. A Feature-based Image Registration Algorithm Using Improved Chain-Code Representation Combined with Invariant Moments, Geoscience and Remote Sensing,

- 1999, 37(5): 2351~2362
- 6 曹闻, 李炳程, 邓子建. 一种基于小波变换的图像配准方法, 测绘通报, 2004, (6): 16~19
- 7 王东峰, 邹谋炎. 基于最大化对齐度的多模态图像自动配准. 电子与信息学报, 2003, 25(6): 755~762
- 8 钟家强, 王润生. 基于边缘的图像配准改进算法. 计算机工程与科学, 2001, 23(6): 25~28
- 9 杨丹, 张小洪. 基于小波多尺度积的边缘检测算法. 计算机科学, 2004, 31(1): 133~135
- 10 Hsieh J-W. Image registration using a new edge-based approach. Computer Vision and Image Understanding, 1997, 67(2): 112~130
- 11 刘贵忠, 鄢双亮. 小波分析及其应用. 西安电子科技大学, 1992
- 12 谯安军, 陈炜, 毛十艺. 一种基于边缘的图像配准方法. 电子与信息学报, 2004, 26(5): 679~684
- 13 张志佳, 黄莎白, 史泽林. 新的基于边缘特征的图像相关匹配算法. 红外与激光工程, 2003, 32(6): 635~638
- 14 Veltkamp R C. Shape matching: similarity measures and algorithms. SMI 2001 International Conference on Shape Modeling and Applications, 2001. 188~197
- 15 Cramariuc B, Shmulevich I, Gabbouj M. A new image similarity measure based on ordinal correlation, 2000 International Conference on Image Processing, 2000. 10~13
- 16 张汗灵, 郝重阳. 基于特征的自动图像配准算法. 视频技术应用与工程, 2003, 9: 80~83

(上接第 219 页)

表 2 近期面向知识发现的网格系统

System	Type	Features
TeraGrid	Network infrastructure	21 TFLOPS computational capacity as a "virtual" system.
InfoGrid	Middleware	Data integration engine. Personalized view of resources for each application domain.
DataCutter	Middleware	Specific for multi-dimensional querying, aggregation and transformation.
Datacentric Grid	Abstract model	Designed for dealing with immovable data. Explicit management of relationships among extracted models.
AdAM	Agent-based framework	Based on repositories of interoperable modules, each specialized for a simple operation.
Discovery Net	KDD middleware	Users compose discovery processes combining Computation and Data services.
Terra Wide Data Mining Testbed	Testbed	Five-node testbed for predictive mining. Generated models are kept up-to-date w. r. t. new data.
Terabyte Challenge Testbed	Testbed	100 nodes organized as a meta-cluster. 2TB storage.
Global Discovery Network	Testbed	Link between the Discovery Net and the Terra Wide Data Mining Testbed.
Knowledge Grid	KDD middleware	domain-independent knowledge discovery environment and services.
Universal Knowledge Grid	KDD middleware	User-centered and domain-independent knowledge discovery and knowledge integration environment and services.

**参考文献**

- 1 Foster I, Kesselman C, Tuecke S. The Anatomy of the Grid: Enabling Scalable Virtual Organizations. Intl J Supercomputer Applications, 2001, 15(3)
- 2 Gruber T. A Translation Approach to Portable Ontology Specifications. Knowledge Acquisitions, 1993, 5: 199~220
- 3 Li Yuhua, Lu Zhengding. Ontology-based universal knowledge grid: Enabling knowledge discovery and integration on the grid. In: Proceedings - 2004 IEEE International Conference on Services Computing, 2004. 557~560
- 4 ISO 9241-11: Guide on Usability (1998)
- 5 Grossman R. Supporting the data mining process with Next Generation Data Mining Systems. <http://www.lac.uic.edu/~grossman/paper/esj-98.htm>. 98 (1998)
- 6 Piatetsky-shapiro G. Knowledge Discovery in Database: 10 years after. SIGKDD explorations, ACM SIGKDD, 2000. 59~61
- 7 Lu Hongjun. Seamless integration of DM with DBMS and Application. PAKDD'01, 2001
- 8 Han J. Data Mining-Current status and Research Directions. Seminar Presentation. <http://db.cs.sfu.ca/sections/publications/slides/slides.html> (2001)
- 9 Haseltine E. Invited Talks: UserCentered Design for KDD. KDD2004 (2004)
- 10 <http://www.kdnuggets.com>, KDNUGGETS Poll Data Mining Applications. htm. (2004)
- 11 Bohanec M, Moyle S, Wetschereck D, et al. A software architecture for data preprocessing using data mining and decision support models (Z). Workshop Integration Aspects of Data Mining, Decision Support and Meta Learning, IDDM 2001, 2001. 13~24
- 12 OWL Web Ontology Language Overview. <http://www.w3.org/>

- 2004/OWL/ (2004)
- 13 Chen Hsinchun, Chung Wingyan, Jie Jennifer, et al. Crime Data Mining: A General Framework and Some Examples. 2004. 50~56
- 14 Hauck R V, Chen Hsinchun. Coplink: A Case of Intelligent Analysis and Knowledge Management. In: Proceedings of the International Conference on Information Systems. 1999. 15~28
- 15 Agrawal R, Gehrke J, Gunopulos D, et al. Automatic subpace clustering of high dimensional data for data mining applications (C). ACM SIGMOD, 1998, 27(2): 95~105
- 16 Xu J J, Chen Hsinchun. Using Shortest Path Algorithms to Identify Criminal Associations. 2002
- 17 Mehta M, Agawal R, Rissanen J. SLIQ: A fast scalable classifier for data Mining (C). In: Proc. of the Fifth Int'l Conference on Extending Database Technology (EDBT), Avignon, France, 1996. 18~32
- 18 Anthony W J, Tung K H, Han Jiawei. Mining Top-n Local Outliers in Large Databases. KDD 2001, ACM, 2001. 293~298
- 19 Kaya M, Alhaji R. Multi-Objective Genetic Algorithm Based Approach for Optimizing Fuzzy Sequential Patterns. In: Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2004), 2004. 396~400
- 20 Han Jiawei, Kamber M. Data Mining Concepts and Techniques. Morgan Kaufmann Publishers, 2001
- 21 Congiustaa A, Pugliesea A, Taliaa D, et al. Designing Grid services for distributed knowledge discovery. Web Intelligence and Agent Systems: An international journal, 2003. 91~104
- 22 Cannataro M, Comito C. A Data Mining Ontology for Grid Programming. In: 1st Int. Workshop on Semantics in Peer-to-Peer and Grid Computing, in conjunction with WWW2003. LNCS (2003)