

基于本体的数据挖掘技术在商务智能中的应用

祁美晶

(大连海事大学 信息科学技术学院, 辽宁 大连 116026)

摘 要 本文首先阐述了本体、数据挖掘技术以及商务智能的含义。介绍了商务智能系统的模块, 也指出了数据挖掘技术应用在商务智能中的重要意义。在数据挖掘中引入本体, 挖掘出的信息更加的全面、准确, 减少了因主观因素所带来的不良后果。结合企业的发展实际, 对基于本体的数据挖掘技术在商务智能中的应用进行了一定的分析, 主要是利用数据挖掘的技术, 在商务智能中发挥出它的真正优势, 为企业做出一定的决策, 提高企业的核心竞争力, 并且促进企业的长远发展。

关键词 本体; 数据挖掘; 商务智能

中图法分类号 TP311 文献标识码 A

Ontology-based Data Mining Technology And Application In Business Intelligence

QI Mei-Jing

(School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China)

Abstract This paper first expounds the definition of ontology, data mining technology and the business intelligence, introducing the module of business intelligence, also pointing out that the application of data mining technology has very important significance in business intelligence. Putting ontology into the data mining can excavate informations that are more all-sided and exact, which can reduce negative effects causing by subjective factor. According to the actual development of the corporations, data mining technology application's case based on ontology is analyzed. It takes advantage of data mining technology to exploit the true advantages to the full, helping enterprise to make decisions, enhancing nuclear competitiveness of companies and stimulate the long-term development of companies.

Keywords Ontology; Data Mining; Business Intelligence

1 引言

随着互联网技术的普及应用, 信息化进入了新的发展阶段, 大大提高了人们的生产及生活水平。各行业逐步实现了信息化的发展道路, 大大提高了各行业的管理水平和经济效益。然而, 随着信息技术的不断推广应用, 企业进入了信息爆炸的时代, 空前的信息量使得企业决策过程越来越复杂, 严重影响了企业对市场的快速响应, 因

此, 在这个全球经济步入信息分析的时代, 管理者需要将大量的、不完全的、有噪音的、随机的实际应用数据转化成有用的商业信息, 问题是如何才能从这些海量信息中发现其中存在的关系和规则, 根据现有的数据来预测未来的发展趋势, 以辅助决策的智能化, 从而带来各种巨大的信息价值。传统的信息处理工具已经不能够满足这一要求。

数据挖掘技术作为一种新兴技术在企业的生

存发展具有重要意义,尤其是在商务智能中的作用更加凸显,因而越来越被广泛应用。随着企业规模的扩大,海量的数据信息增加了企业管理者决策的难度。在新时期,在信息处理过程中,管理者需要一种自动分析数据、自动发现和描述数据中隐含的商业发展趋势、对数据进行更高层次分析的方法。从而准确预测出企业发展趋势,促进企业良性发展。

2 本体与数据挖掘技术

2.1 本体

本体(Ontology)的概念最早来源于哲学,后来随着科学技术的快速发展,本体被赋予了新的含义。最早将本体引入信息应用领域的是人工智能界的研究人员。最早给出定义的是Neches等人。他们对本体的定义是“给出构成相关领域词汇的基本术语和关系,以及利用这些术语和关系构成的规定这些词汇外延的规则的定义”。再后来,被引用最广泛的定义是由Gruber提出的,“本体是概念化的明确的规范说明”。本体的目标是在相关领域中获取知识,对该领域的知识提供共同的理解,将该领域内的共同认可的词汇进行确定,并从不同层次的形式化模式上给出这些词汇和词汇之间相互关系的明确定义。

构造本体的目的是为了分析澄清领域知识的结构。为知识表示打好基础。它的作用可总结为:通讯(communication)、互操作(inter-operability)和系统工程(systems engineering)。在构造过程中,本体可以用下面的公式来形象的表示:

本体=概念(Concept)+属性(Property)+公理(Axiom)+取值(Value)+名义(Nominal)

五种经典的构造本体的方法:Uschold和King方法、Gruninger和Fox方法、Bernerlas方法、基于SENSUS方法、MethOntology方法。

事实上,对于构建一个真正的本体来说,并没有一种唯一的途径或者方法。对于任何专业学科领域,都不存在某一种唯一适合的途径或者模式,可能存在好几种方法都是可行的。最佳的方法依赖于本体的构建者所采用的应用软件以及可以预见的扩展功能。同时,本体的开发和完善是一个反复叠加的过程,不会一蹴而就。无论从客观世界具体对象的角度出发,还是从抽象的角度出发,本体中概念的设计应贴近专业领域、客观

对象和对象间的关系法则。

在智能领域,本体则是数字模型与资源共享的体现,它是一种形式上的规范说明,从而更加明确、具体阐述了概念与概念之间较为复杂的关系,由于本体有着较强且健全的约束力,因而在本体思想下,各种数据集成和相关软件的应用更加容易实现[1]。在现代网络知识中融入本体知识,能够使知识、信息更加对象化,利于知识集成的实现。同时,通过对知识对象之间的相互匹配,能够使匹配知识对象之间的关系更加明确,利于清晰的对各种关系进行有效描述。利用各种知识关系之间的属性,可以使用户在短时间内获取到自己需要的知识信息内容,从而避免了在盲目的状态下,获取大量不必要的信息数据,从而节省了工作时间,提高了工作效率,并且较为清晰、准确的数据更加有利于用户进行决策。

2.2 数据挖掘技术

随着数据库技术的不断发展,数据库管理系统也得到了广泛的应用,当数据库中的存储数据量急剧的增大时,数据理解和数据产生之间就会出现越来越大的距离。而数据挖掘(Data Mining)就是为解决这一矛盾的一种新型数据分析技术。数据挖掘是一个从数据集中识别出有效的、新颖的、潜在有用的,以及最终可理解的模式的非平凡过程。数据挖掘主要的目的是从大型的数据库中提取隐藏的预测性信息。它可以发掘数据间潜在的模式,找出企业经营者可能忽视的信息,以理解和观察的形式反映给用户,为企业做出有预见性的、基于知识的决策参考意见[2]。

数据挖掘主要是通过对海量的数据信息进行有目的性的提取、分拣、归类,挖掘隐含的有用信息。数据挖掘的任务需要利用聚类和分类作为基本分析手段,将大规模异种类别的数据按属性相似划分为子集,有利于减少数据处理的规模,简化分析和建造模型的复杂性。[3]

通常数据挖掘需要有数据清理、数据变换、数据挖掘实施过程、模式评估和知识表示等八个步骤。

(1)信息收集:根据确定的数据分析对象抽象出在数据分析中所需要的特征信息,然后选择合适的信息收集方法,将收集到的信息存入数据库。对于海量的数据,选择一个合适的数据存储和管理的数据仓库是至关重要的。

(2)数据集成:把不同来源、格式、特点性质

的数据在逻辑上或物理上有机地集中，从而为企业提供全面的数据共享。

(3)数据规约:执行多数的数据挖掘算法即使在少量数据上也需要很长的时间，而做商业运营数据挖掘时往往数据量非常大。数据规约技术可以用来得到数据集的规约表示，它小得多，但仍然接近于保持原数据的完整性，并且规约后执行数据挖掘结果与规约前执行结果相同或者几乎相同。

(4)数据清理:在数据库中的数据有一些是不完整的，含噪声的，并且是不一致的，因此需要进行数据清理，将完整、正确、一致的数据信息存入数据仓库中。

(5)数据变换:通过平滑聚集，数据概化，规范化等方式将数据转换成适用于数据挖掘的形式。对于有些实数型数据，通过概念分层和数据的离散化来转换数据也是重要的一步。

(6)数据挖掘过程:根据数据仓库中的数据信息，选择合适的分析工具，应用统计方法、事例推理、决策树、规则推理、模糊集等方法处理信息，得出有用的分析信息。

(7)模式评估:从商业的角度，由行业专家来验证数据挖掘结果的正确性。

(8)知识表示:将数据挖掘所得到的分析信息以可视化的方式呈现给用户，或作为新的知识存放在知识库中，供其他应用程序使用。

数据挖掘过程是一个反复循环的过程，每一个步骤如果没有达到预期的目标，都需要回到前面的步骤，重新调整并执行。不是每件数据挖掘的工作都需要这里列出的每一步，例如在某个工作中不存在多个数据源的时候，步骤2便可以省略。步骤3,4和5又合称数据预处理。在数据挖掘中，至少60%的费用可能要花在步骤1的阶段，而至少60%以上的精力和时间是花在数据预处理上的。

3 商务智能

商务智能(Business Intelligence)是1989年由Garther Group的Howard Dresner首次提出，他将商务智能定义为一类由数据仓库、查询报表、数据分析、数据挖掘、数据备份和恢复等部分组成，以帮助企业决策为目的的技术及其应用。该技术收集、汇总了与商务活动有关的各种数据，

将其集成到数据仓库中，采用联机分析技术对商务活动进行实时的监控、分析，便于及时采取有效的商务决策，提升商务活动的绩效。应用数据挖掘技术对描述商务活动的数据进行挖掘，以获取有效的商务信息，从中提取商务知识。它也是一种汇集了机器学习、信息模式识别、信息学、数据库、统计学，以现代管理理论为指导，信息技术为支撑核心的集成系统；描述了一系列的概念和方法，通过应用基于事实的支持系统，来辅助商业决策的制定。通过BI技术可以提高企业效益，建立忠实的顾客群，并且能增进企业效益以及帮助企业做出明智的决策。商务智能的体系结构如图1所示。

商务分析：绩效管理、客户管理、供应链管理
OLAP、数据挖掘
数据仓库

图1 商务智能体系结构
商务智能系统模块的内容如图2所示。

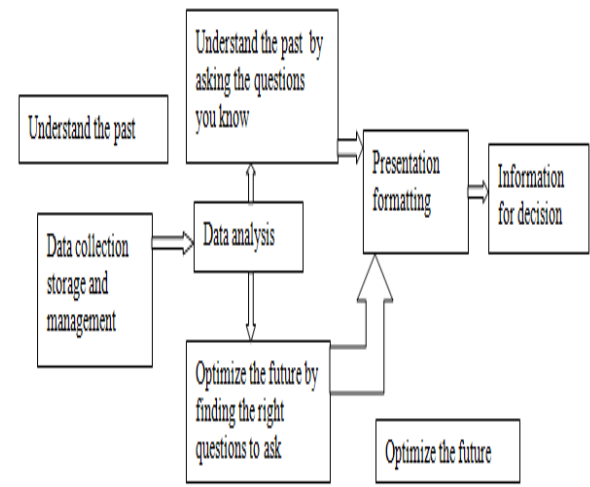


图2 商务智能系统模块
商务智能工作的过程主要是实现对相关数据信息流动与储存的过程，商务智能系统通常主要有以下几个模块组成：

- (1) 数据获取
在这个过程中，主要实现对源数据的采集、筛选、整理、转换以及存储。
- (2) 数据管理
该模块利用数据挖掘技术，主要负责数据仓库的内部维护和管理。主要进行数据存储的组织、数据的维护、数据的分发、数据安全、数据提取、

数据清晰、数据转换以及数据维护。通过数据库的管理,能够对众多的数据信息进行提取、净化、过滤以及数据标准化等。

(3) 数据分析

在这个阶段,商务智能系统真正实现了智能化,该阶段主要利用联机分析处理和数据挖掘技术,对从数据仓库中提取的数据进行汇总和多维分析,挖掘出数据背后隐藏的知识。

(4) 信息呈现

这个模块是将以上数据分析所得到的决策知识展现在用户或者是企业管理者面前,支持管理和决策。

数据挖掘的过程如图 3 所示:

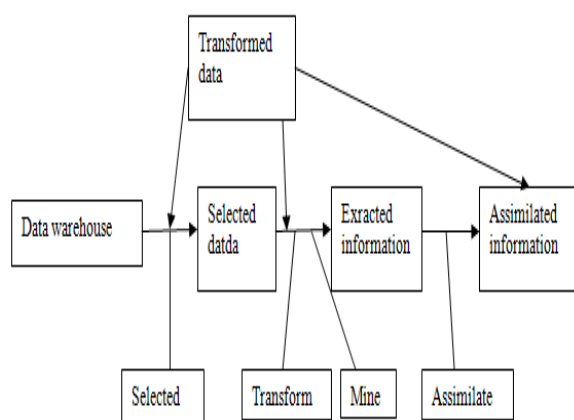


图 3 数据挖掘过程

4 数据挖掘技术在商务智能中应用

4.1 数据挖掘技术与商务智能系统

在企业中,随着业务规模的扩大,数据库中储存的信息数量在不断的增加,因此出现了数据挖掘技术。这种新型的数据分析技术能够有效解决众多数据之间出现的矛盾。被应用的领域也很广泛。它可以挖掘出潜在的、有用的信息,从而帮助企业的决策与发展提供了有预见性的参考意见。利于企业的发展。

企业想要在运作的过程中将大量流动的商业数据能够发挥出更大的作用和功效,企业在发展运行中应用的商务智能系统就需要采用数据挖掘技术,进而才能使各种商务信息得到有效的发挥。总之,数据挖掘技术在商务智能系统中起到了很重要的作用。它在商务智能系统中的位置如图 5 所示:

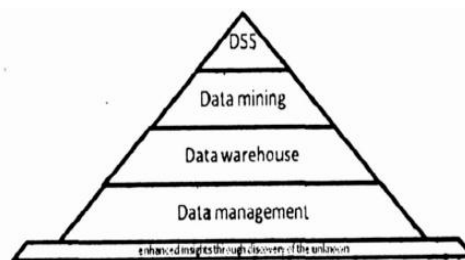


图 5 数据挖掘在商务智能系统中的位置

4.2 高层次规则处理的优势

与低层次数据挖掘规则进行比较,可以清楚的发现,在高层次规则引导下进行的数据挖掘具有以下优势:

(1) 高层次规则能够清晰准确的对所提供的数据进行概括和总结。

(2) 在高层次的数据提取中使用了具有相关性的搜索方略,因此与低层次数据挖掘相比,数据规则的数量就会减少。当低层概念转化为高层概念时,就会得到少数的规则,同样的,单个高层次规则也会替代掉形式与内容较为形似的低层次规则。

4.3 数据挖掘技术的影响因素

在当前各大企业的知识管理中,普遍都存在一定的问題,影响了数据挖掘的过程:

(1) 很多的企业在管理过程中,由于缺乏较为统一的知识管理模型,因而很容易使系统对同一个类型的知识描述出现不相同的情况,影响用户对这一知识的理解会有偏差,不能准确领会知识的真正含义,以及不能真正实现知识信息资源的有效共享。

(2) 在企业管理中,尤其是在知识储存模式,由于对知识储存没有做出严格规定,缺乏统一的知识存储模式,这就造了各个知识的存储形式各不相同,对用户知识使用造成了影响,妨碍了人们的知识交流,没有真正地发挥出知识资源的优势。

(3) 在知识的检索和查全率、查准率性上有一定的问题,对检索到的知识缺少语义和山下文的支持。

4.4 基于本体的数据挖掘技术

为了更好地解决上述问題对商务智能系统造成的影响,引入了本体。根据企业的实际发展情况,可以建立某个领域的本体,把本体作为出发点引导数据库技术的应用,对数据挖掘的过程进行相应的完善,进而可以挖掘出更有效的数据信

息。因此，本文提出了一种新的数据挖掘方法，该方法是建立在本体基础之上，数据挖掘方法的详细描述如图4所示：

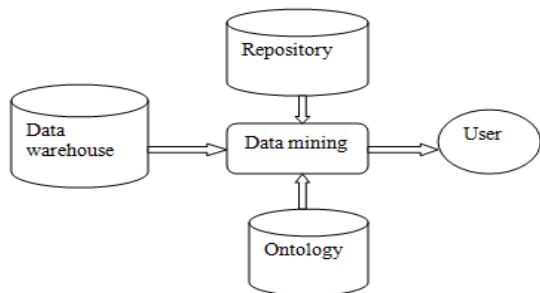


图4 基于本体的数据挖掘

利用本体知识或者相关背景性的知识，在高层次建立数据挖掘，相应的会产生高层次或多层次规则。在特殊的情况下，会产生具有语义性意义的一些规则，这些规则通常是由具有抽象性的高层次概念组成。此外，该商务智能系统可以实现数据的自动化挖掘，通过本体对数据进行预处理以及后处理。

5 结束语

基于本体的数据挖掘技术对各个领域的发展具有重大的意义，也就意味着它具有广阔的应用前景。基于本体概念的数据挖掘，能在多层次泛化基础上引导挖掘过程，并裁剪规则空间，能产生高层次，甚至是具有语义的规则上产生挖掘结果。在今后的工作中要不断地完善本题库的内容，建立可表示、可操作的。粒度足够细的本体知识库，实现智能的人机交互、计算机系统之间的互操作和计算机系统知识重用，其次是将知识库与数据库的结合，使该系统能够更好地发现更有价值的信息。

我们需要立足企业发展实际，认真分析数据库技术的特点，加强对数据库挖掘技术的研究与应用，实现人机交互智能化，促进企业商务智能系统的设计，提高企业决策的科学性，从而能够促进企业的长远发展。

参考文献

- [1] 宋远芳, 卢明月, 赵海燕. 基于本体的数据挖掘技术在商务智能中的应用[J]. 计算机技术与发展, 2013, 12(01):101-102.
- [2] 韩家伟, 坎伯, 范明. 数据挖掘概念与技术[M]. 北京: 机械工业出版社, 2001.
- [3] 中华论文网. 知识发现与数据挖掘 [EB/OL]. 2007.

<http://www.51ky.com>.

- [4] 李健康, 张春辉, 孟晓花. 本体研究及其应用进展[J]. 图书馆论坛, 2012, 13(12):80-86.
- [5] 吴晓蕾. 基于 Ontology 的内容分析法的研究[D]. 南京: 南京理工大学, 2005.
- [6] 穆军. 基于本体的数据挖掘技术应用于商务智能中的实际应用, 开发应用[J]. 微型电脑应用, 2014, 30(12):48-50.
- [7] 卢美律, 谢明亮, 李明明. 数据库里的知识发现[J]. 中国人民大学图书馆复印报刊资料, 2012, 13(02):127-130.
- [8] Kantardzic M. 数据挖掘概念、模型、方法和算法[M]. 北京: 清华大学出版社, 2011, 09(12):13-14.
- [9] 格罗斯 R. 数据挖掘——构筑企业竞争优势[M]. 侯迪, 宋擒豹译. 西安: 西安交通大学出版社, 2001.
- [10] 李小平, 张明亮, 王海青. 数据采掘技术回顾[J]. 小型微型计算机系统, 2011, 15(04):74-81.
- [11] 刘平峰. 基于知识网格的电子商务智能推荐理论方法研究[J]. 武汉理工大学学报, 2012, 10(09):17-18.
- [12] M Uschold, M Gruninger. Ontologies :Principles, Methods, and applications. Know ledge Engineering Review, 1996, 11(02): 93-155