

基于数据仓库与数据挖掘技术的包装信息处理方法研究

金艳, 纪钢

(重庆理工大学, 重庆 400054)

摘要: 针对包装信息的特点, 阐述了包装信息数据仓库的建立方法, 把各种包装信息内部数据和外部数据进行有效集成, 并同时阐述了在包装信息数据仓库下进行包装信息数据挖掘的方法, 实现了将包装信息中的数据挖掘技术集成到复杂的包装信息技术应用环境中, 为各层次的研究决策、分析人员服务。

关键词: 数据仓库; 数据挖掘; 包装信息; 包装设计

中图分类号: TB488 **文献标识码:** A **文章编号:** 1001-3563(2012)17-0148-03

Study of Packaging Information Processing Method Based on Data Warehouse and Data Mining Technology

JIN Yan, JI Gang

(Chongqing University of Technology, Chongqing 400054, China)

Abstract: The packaging information data warehouse building method was introduced, which effectively integrated all kinds of internal data and external data of packaging information according to the characteristics of packaging information. The data mining method of packaging information under the condition of packaging information data warehouse was introduced, which integrated data mining technology into the application environment of complex packaging information technology. The purpose was to provide reference for all levels of decision-maker and analyzer.

Key words: data warehouse; data mining; packaging information; packaging design

由于目前包装信息所涉及的面较广、信息量大, 缺乏对包装信息及数据的存储及归纳。在进行各种包装设计时, 还必须提供大量的与包装信息相关的各种资料, 包括包装设计标准、包装设计的参考资料, 要获取这些信息, 需要花费大量的人力、物力和时间, 使得各种包装设计在很大程度上不是依靠信息的共享, 而是依赖于包装设计人员自身技术储备、以往的经验以及对包装产品所进行的一些试验所得到的数据结果^[1]。另外, 在一般包装信息管理中, 由于需要存储包装物从生产过程到产品的分类管理, 再到再生利用各环节的包装信息数据时, 就会发现这种数据没有进行分析处理, 其表示的结构性较差, 用一般的关系型数据库进行存储、包装信息的查询、包装信息的分析处理中, 其包装信息的可用性、准确性在实际使用中, 将会产生一些问题。由于包装材料的类型及分类较为复杂, 所以, 在包装信息的使用中, 它们的分类、实

际有用信息的获取, 在一般信息处理中, 就会产生一定的问题, 有必要引入数据仓库与数据挖掘技术进行包装信息处理方法的应用研究^[2]。

1 包装信息处理技术

1.1 包装信息存储技术

在普通包装信息数据库的存储中, 由于存储的单一性, 不能较为完整地进行包装信息的筛选及分析, 特别是在并行数据的广泛应用、数值化信息的快速增长、信息的极度膨胀下, 引入数据仓库(DW)在包装信息的处理中已显得尤为重要。包装信息数据仓库应是一个面向包装信息主题的、相对稳定的、集成的、反映包装信息历史变化的数据集合, 它是在各种包装数据库存在的情况下, 为了进一步挖掘包装信息有用的数据资源、为了包装信息决策需要而产生的。根据目

收稿日期: 2012-07-08

作者简介: 金艳(1978—), 女, 四川资中人, 重庆理工大学讲师, 主要研究方向为数据库、计算机应用。

前需要的决策目标,将存储在各个数据库中的包装信息数据通过一定的选择,转换为一种新的信息存储格式,其目的是解决那些具有大量的包装信息数据,而在应用中又无法获取有用信息的问题^[3-4]。所以,包装信息数据仓库需要进行的工作有:

1)数据源。指包装信息数据仓库中的数据,可来源于多个、多种形式下的包装信息数据源,其中可包括各种数据文件、各种外部文档资料等。

2)仓库管理。在包装信息数据仓库信息组成确定后,需要建立相关的包装信息数据模型,然后确定在源数据仓库中的包装信息数据的抽取转换等实施处理方法,最后确定包装信息数据仓库的物理存储结构及数据库管理系统的功能。

3)分析工具。在包装信息数据仓库下,建立用于解决在实际分析、决策中,其系统所需的各种查询、信息检索等分析工具。

1.2 包装信息数据挖掘技术

随着包装信息数据量急剧增大,如何从大量的包装信息数据背后所隐藏的许多重要的信息中,抽取有用信息,是包装信息数据挖掘技术所要解决的问题。包装信息数据挖掘是从大量的、有噪声的、不完全的、模糊的、随机的包装信息数据中提取隐含在包装信息数据中的、事先不知道但又是潜在有用的包装信息和包装相关知识的一种处理技术^[5-7]。目前,在包装信息数据挖掘技术处理中,主要算法及方法有:

1)遗传算法。在包装信息数据挖掘中,由于遗传算法是基于生物自然选择与遗传机理的一种随机搜索算法,具有的隐含并行性、易于和其他模型结合。

2)神经网络方法。由于该方法具有自组织自适应性、有良好的鲁棒性、并行处理、高度容错和分布存储等特性,所以,在包装信息数据挖掘中,非常适合解决包装信息数据挖掘的问题。

3)决策树方法。是一种预测模型的处理算法,它通过将大量数据有目的地进行分类,从中找到信息集中有价值的、潜在的信息。所以,在包装信息信息处理中,由于描述简单、分类速度快,可在包装信息数据仓库中进行大规模的数据处理。

4)统计分析方法。主要利用统计学知识对包装信息数据库中的信息进行分析,解决大量包装信息数据中的最大值、最小值、总和、平均值等统计信息的计算^[5-7]。

2 数据仓库与数据挖掘技术在包装信息处理方法中的应用

2.1 数据挖掘与数据仓库在包装信息处理方法中的关系

在包装信息数据挖掘中,与包装信息数据仓库紧密联系在一起时,将在包装信息数据仓库中获取所需要的包装信息及包装知识内容。所以,在包装设计、包装信息的应用中,首先要建立包装信息数据仓库,该数据仓库的联机分析功能可为包装信息数据挖掘提供相应的操作平台。在包装信息数据仓库的构造过程中可以围绕包装信息数据仓库组建,进行包装信息数据存取、数据集成、数据合并、异种数据库的转换、ODBC/OLE DB 的连接、Web 访问等各种数据分析和数据分析等基础工作。这样,在进行包装信息数据挖掘与包装信息数据仓库的有效联结时,将增加包装信息数据挖掘的联机挖掘功能,充分利用包装信息数据仓库的 OLAP 与各种包装信息数据挖掘方法及工具的联结,为包装信息数据挖掘选择合适的数据挖掘方法及工具。

2.2 包装信息数据仓库的处理方法

包装信息数据挖掘首先要建立在基于包装信息数据仓库的一种有效的处理算法,从包装信息数据中抽取处理模式。对于包装信息数据挖掘的应用不仅需要建立良好的、有效算法处理模型,而且更重要的是要解决如何将包装信息中的数据挖掘技术集成到复杂的包装信息技术应用环境中。由于包装信息种类繁多,在获取大量包装信息数据之后,需要寻找包装材料的种类、包装外观设计的方法、包装应用的使用范围,并进行预测、预报。在进行包装信息数据仓库处理中,主要进行的工作有:构建包装信息数据仓库的基本理论和方法;实施对包装信息数据仓库的有效管理及维护;确定如何快速收集和有效存取包装数据的有效方法及处理技术;建立和完善包装信息元数据字典,为数据挖掘提供底层的基础数据信息;在数据信息的分析处理中,建立如何从包装信息数据仓库中的综合数据中恢复原始数据的功能。

2.3 数据挖掘技术处理方法

在对包装信息中的选材、设计、应用环境等数据信息进行数据挖掘,实质就是知识和规则的发现过程。在知识发现系统中,就是从大量的包装信息数据中筛选信息,发现有价值的知识和规则。从包装信息

数据仓库中选择的数据在知识发现引擎里进行处理,在引擎中提供各种挖掘算法,从中生成有潜力的、有效的应用价值知识。在基于包装信息数据仓库的知识发现系统结构中建立相关包装信息知识发现系统,就是要在包装信息处理中,进行包装信息数据挖掘研究与开发包装信息数据挖掘的理论与技术,设计高效的、有针对性的包装信息算法挖掘规则以及高效算法规则的更新、维护和管理功能。所以,在包装信息数据挖掘技术处理中,需要进行的处理方法有:

1) 关联分析。就是在包装信息数据库中解决重要的、2个或2个以上变量的取值之间存在某种规律性、可被发现的知识,在处理中可进行简单关联、时序关联和因果关联。其目的是找出包装信息数据库中隐藏的关联网,解决包装信息中各种包装信息数据之间的关联度。

2) 聚类分析。首先把包装信息数据按照相似性归纳成若干类别,同一类中的包装信息数据彼此相似,不同类中的包装信息数据相异,然后,通过聚类分析与处理,建立包装信息类别的宏观概念,发现包装信息数据的分布模式,以及可能的包装信息数据属性之间的相互关系。

3) 分类。就是在包装信息中,找出一个包装信息类别的概念描述,以代表该类包装信息数据的整体信息,即该类的内涵描述,并用这种描述来构造包装信息数据模型。

4) 预测。首先利用包装信息的历史数据,从中找出包装信息数据的变化规律,并建立相关包装信息处理模型,然后由此模型对未来包装信息数据的种类及特征进行预测,用预测方差来度量它的准确性。

3 结语

包装信息数据仓库是包装信息处理中包装信息数据库发展和应用的延伸,是在海量的包装信息数据库中提取潜在的、有价值的包装信息的一个实体,而包装信息数据挖掘是包装信息知识发现的一个重要

的处理技术,把数据仓库和数据挖掘技术应用到包装信息处理研究中,可以提高包装信息处理、有用信息的使用,发现更多而有实际应用价值的信息,使宝贵的包装信息数据资源在包装设计、包装信息处理中得到有效利用。

参考文献:

- [1] 朱文求. 数据仓库技术在包装产品网络营销中的应用探讨[J]. 包装工程, 2002, 23(5): 21-22.
ZHU Wen-qiu. An Explore on Date Warehouse Technology Applied in Web Market for Packaging Products[J]. Packaging Engineering, 2002, 23(5): 21-22.
- [2] 任其昌, 袁友伟. 数据挖掘在包装产品网络营销中的应用[J]. 包装工程, 2001, 22(3): 40-41.
REN Qi-chang, YUAN You-wei. Applying Data Mining to Web Marketing for Packaging Products[J]. Packaging Engineering, 2001, 22(3): 40-41.
- [3] 萨师煊, 王珊. 数据库系统概论[M]. 北京: 高等教育出版社, 2009.
SA Shi-xuan, WANG Shan. Database System Introduction[M]. Beijing: China Higher Education Press, 2009.
- [4] 韩家炜, 坎伯. 数据挖掘概念与技术[M]. 北京: 机械工业出版社, 2007.
HAN Jia-wei, KAMBER M. Data Mining: Concepts and Techniques[M]. Beijing: China Machine Press, 2007.
- [5] AGOSTA L. The Essential Guide to Data Warehousing [M]. London: Prentice Hall PTR, 1999.
- [6] GIOVINAZZO W A. Object-oriented Data Warehouse Design[M]. London: Prentice Hall Press, 2000.
- [7] 沈兆阳. 数据仓库与 Analysis Services[M]. 北京: 清华大学出版社, 2001.
SHEN Zhao-yang. Data Warehouse and Analysis Services[M]. Beijing: Tsinghua University Press, 2001.
- [8] 陈京民. 数据仓库与数据挖掘技术[M]. 北京: 电子工业出版社, 2002.
CHEN Jing-min. Data Warehouse and Data Mining Technology [M]. Beijing: Electronic Industry Press, 2002.

(上接第 143 页)

- [6] BASTIDAS D M. Anti-Corrosion Methods and Materials [J]. Volatile Corrosion Inhibitors, 2005, 52(2): 71-77.
- [7] 梁成浩. 防锈溶剂型可剥性塑料的研究现状与展望[J]. 腐蚀科学与防护技术, 2009(11): 546-548.

LIANG Cheng-hao. Retrospect and Prospect of Antirust Strippable Plastic[J]. Corrosion Science and Protection Technology, 2009(11): 546-548.