

数据仓库与数据挖掘研究进展

王丽敬

中国人民大学 北京 100085

摘要: 数据仓库简称数据库, 数据库, 在近年来发展的很迅速, 也融入了大学生的考试当中, 数据仓库技术指的是将所有的数据整理起来, 统一放在一个中央仓库中, 将所有的数据, 加以理解并通过公式和模型的方式计算出所有的数据结果, 并给用户呈现所有的看法与决策, 数据仓库从整体上来看是具有历史性的, 他是在近年来一点点不断发展起来的, 因此, 他发展史不长, 只是在面对历史数据时, 具有一个参考性价值, 并且还有只读性的特点, 指的是在一个数据中, 指定某一个数据可读, 因此, 其根本目的在于给企业做出决策的基础和依据, 数据挖掘简单来说, 就是将所有的数据整合出来, 找到并整合出来。

关键词: 数据仓库; 数据挖掘; 研究进展

Research progress of data warehouse and data mining

Lijing Wang

Renmin University of China, Beijing 100085

Abstract: data warehouse is called database for short. Database has developed rapidly in recent years and has been integrated into the examination of college students. Data warehouse technology refers to sorting all the data into a central warehouse, understanding all the data, calculating all the data results through formulas and models, and presenting all the views and decisions to users, On the whole, data warehouse is historical. It has been developing little by little in recent years. Therefore, its development history is not long. It only has a reference value when facing historical data, and it also has the characteristics of read-only. It means that a certain data is specified in a data to be readable. Therefore, its fundamental purpose is to provide the basis and basis for enterprises to make decisions. In brief, data mining, Is to integrate all the data, find and integrate them.

Keywords: data warehouse; Data mining; Research progress

引言:

在当代社会数据挖掘技术也发展的越来越迅速, 人们也越来越喜欢用挖掘技术和数据仓库技术来整合数据, 一旦数据挖掘过程中发现这些数据有可以利用的价值, 数据仓库技术就会将这些数据整合起来, 加以利用并做出相应的决策判断, 尤其是在于企业的生产经营当中, 某些财务数据一起整合起来, 然后再加以判断, 可以更有利于该企业生产经营的发展, 从目前来看, 我国的数据库还处于发展, 不完全阶段, 在发展数据库时也有很多不健全, 不完全的因素, 例如说, 在数据仓库处理过程中, 我们缺乏该专业的人才, 没有真正的行业特

点体现应用技术, 但是不管怎么说数据仓库在我国发展如此迅猛, 因此, 我国在发展数据仓库和数据挖掘过程中, 还应该有很多学习和改善的内容, 我国应该不断建立健全数据仓库和数据挖掘的各项策略研究。

现代社会的企业不仅仅追求经济上面的发展, 还追求在决策中积累经验, 因此, 在决策中积累经验, 就应该采用到数据整合以及分析推理的方法, 这样才能让企业做出的决策更加高效, 企业发展的也会, 稳定健全的发展。企业如果不能准确的得到消息并且分析数据, 就没有办法精准的做出相应的决策, 在这一关键过程中, 建立数据库, 就成了解决办法的途径, 当企业建立了数据仓库之后, 可以从企业的信息中筛选出自己想要的, 数据仓库不仅仅是记录一个数据, 而是记录多品种多方面的数据, 并且将推演出正确的决策方式, 通过数据仓

作者简介: 王丽敬 (1990.08), 女, 汉, 河南, 本科, 高级研发, 在一家新闻资讯公司从事研发工作。

库，我们不仅仅可以看到关系网的变化，也可以看到，每一个公司所经营产品的财务信息，它可以将企业的数据库汇集总结在一个总的数据库中，并将其转化成一个小的数据库，举例一个公司的财务信息，我们要想看到一个公司的财务信息，就必须知道其生产经营所带来的销售利润和生产经营收入以及其成本有哪些方面，在了解了这些问题之后，我们可以在其销售利润的数据库中看到每一年的销售利润有多少，生产经营收入的数据库中看，其生产经营收入，在每一个数据库中，我们可以推演该企业在明年可以取得多少利润，从而做出相应的决策，该公司该稳定发展还是激进发展^[1]。

一、数据仓库与数据挖掘

(一) 数据仓库设计概述

对比内容	数据库系统设计	数据仓库系统设计
面向的处理类型	面向应用	面向分析
应用需求	比较明确	不太明确
系统设计的目标	事务处理的并发性、安全行、高效性	保证数据的四个特征和全局一致性
数据来源	业务操作员的输入	业务系统
系统设计的方法	需求驱动	数据驱动

(二) 基本概念

数据仓库用通俗的话来讲，就是将所有的数据整合起来，放在电脑上一个系统的仓库之中，一些数据放到仓库之中之后，是一直都不会变化的，也会一直保存数据都分有自己的主题，具有相对性和独立性，而数据仓库系统则是，一个一个数据仓库结合起来，总和就成了数据仓库系统，数据仓库系统的主要功能则是将一个个数据仓库之中的内容整合并加以利用，分析得出自己的结果判断，使得企业可以做出正确的决策，数据挖掘则是在所有的数据当中找到自己想要的数据库，因此，在数据挖掘的过程中，数据仓库成了基础，在数据仓库中，数据挖掘又是必然的，两者不可或缺。

二、数据仓库的结构、功能

(一) 数据仓库的基本结构

数据仓库中的数据是怎么保存的？我们应该根据数据的什么类型来进行分析处理，应该分为什么层次来进行处理，因此，数据仓库的基本结构又分为很多个部分：

1. 历史性详细数据层：用通俗的意义讲，该数据库中储存的数据通常为很早之前的数据，一般来说，为5到10年前的数据，甚至是十年以上的数据，这些数据适用于对比，对比现在发展和对未来的预测，这种数据库只用于进行分类储存，举例，一个公司在财务方面进行

了5到10年的数据保存，我们就可以知道该公司在5到10年内发展的情况，并且可以预测，该公司未来发展是走弱或是走强。

2. 当前详细数据层：当前，详细数据指的是线下必须要记录的所有数据，但是他又有时间的推荐，根据时间，当前的数据总会成为历史的数据，五年之后，当前的数据也变成了历史性的数据，因此，当前数据只是为了了解目前公司的生产经营状况，只是了解现在公司生产经营到底好不好^[2]。

3. 不同程序的归纳总结信息层：不同程序的归纳指的是在数据库当中，我们可以建立不一样的数据库，例如说，根据财务报表，我们可以建立及周报，月报或是，年报的统计，这些数据，只是每天记录公司生产经营状况，不能对未来的预测形成多大影响。

4. 专业信息分析层：专业信息指的是我们通过数据库对信息内容加以分析，通过不同的方法统计数据并加以利用，形成一个专业结果预测其未来走向

5. 仓库结构信息：仓库结构在通常意义上讲，仓库的内部会存在许多检索，查询，索引的内容，我们应该在每个数据库当中都做出较大的索引，这样才可以有利于其查询，如若是在整理数据仓库中，所有数据时，我们应该观察其数据的访问量，对于访问量较少的数据，不必做太多索引，对于访问量较多的数据，应该将其建立索引，这样让其有目标的查看自己想要的数据库。

(二) 数据仓库的功能特点

介绍完数据库的分类以及基本信息之后，我们将介绍数据库的功能特点，数据仓库的主要，功能是为了让企业做出更好的决策，整合所有的数据信息，可以在利用索引的情况下，随时获取自己想要的数据库信息，在经过这些数据信息将企业的所有经营生产状况全部列出并，演算出该公司日后的生产经营状况，当然，数据库中有很多表格，因此，我们必须每条做好分类。具体的功能特点有：

1. 面向对象性。在数据仓库中建立每一个数据库，都是为了将问题得到有效的解决，因此，我们在建立数据库时，一定要真有针对性的建立数据库，将不同种类，不同角度的数据库进行数据加工整理，并且保存下来

2. 数据历史性。数据仓库具有历史性，因此，数据也具有历史性，数据仓库中可以保存很久的数据，因此，数据必须具有，时间的标注，我们才可以更加准确地做出数据之间的比较，并且对其进行预测。

3. 数据集成性。数据无论是怎么得来的，在进入数

据仓库之后，都会有统一的标注，因此，数据仓库中所有数据都是集成一体的。

4.数据只读性。一旦数据录入到数据仓库之中，数据仓库将只对其提供只读服务，在录到数据库当中，之后数据将不可以进行改动，只能刷新数据，因此，在数据录入数据仓库当中时，应该把握数据的准确性。

5.操作集合性。来自不同地方，不同网站的数据都可以将其录入自己的数据库当中，这样子也可以把自己的数据库建立的更加完整。

三、数据挖掘技术

数据挖掘指的是在自己建立的总的数据库当中，我们可以提取到自己需要的消息，主要用于将进行预测信息的技术数据挖掘，是一种专业挖掘信息以及预测的工具，它能够发现很多潜在信息，例如他可以，在一个公司的财务报表中，提取出其生产经营状况所能展现的利润，生产收入和生产成本等，这样就可以让企业的管理人员做出预测性的决策，当然，数据挖掘技术也不单单只有一种传统的技术，通常是采取假设的前提下，然后对所有的数据进行分析验证，这个假设是正确的，还是错误的，因此，数据挖掘现在只能对所有的数据进行整理，并且做出预测性的分析，从而让公司做出正确的相应决策，因此，将其业务越做越好，帮助决策者更好地掌握市场的策略^[3]。

四、联机分析技术

联机分析技术用通俗的话来讲，就是数据仓库在，会产生各种演算的过程，因此在这些演算的过程当中，我们就应该产生一种新的技术，这种新的技术就叫联机分析技术，总体上来讲，联机分析技术，主要作用就是应对于出现大量的数据仓库中数据演算的操作，从而产生的技术，它可以分为两种类型，一种是针对于数据仓库产生的系统，这种系统主要是对于数据仓库中的模式进行，加工处理，并且通过其数据仓库中索引进行查询，并且进行分析操作，目前在很多大型厂商中加以利用，主要原因在于，大型厂商需要记录多种商品的价格，进货量等，这样子运用连接分析技术，可以增加其寻找商品的快速性，当然，建立数据仓库，联机分析技术，也应该加入，索引这样更方便，对数据的查找，第二种极是建立多维模型，由于有些数据不单单只是需要查询那么简单，那么我们就需要建立模型查看系统，建立模型主要研究的问题是储存空间的问题，数据仓库的储存空间是非常大的，因此，我们必须建立多维模型，才能促进数据仓库系统的正常有效发展，还可以维持数据仓库

系统的有效运行。

五、一种基于DW的DMS结构

数据仓库有利于方便等特点，根据这些特点，我们可以提出一种结构框架，这些结构框架的类型基本概念如下：

1.用户在查询过程中，总会碰到不知道该从哪个窗口进行查询，因此，我们可以将查询界面分为分类以及解释两个部分，它的作用就是将数据挖掘更加准确，可以让用户更加快速的找到自己想要的的数据，这样子查询也可以处理的，非常方便。

2.在查询协同过程中，查询协同的作用，就是将数据仓库中的所有数据利用数据挖掘工具，将所有的数据对接起来，对接起来，通过同一个窗口进行查询。

3.数据仓库管理系统它直接负责对数据仓库进行管理，并完成对各种异构分布数据源中数据的提取工作，以最大限度屏蔽各异构数据源对系统的影响。

4.数据管理系统中还包括知识库的管理，这是库管理应当对录入的所有知识，进行分类处理，并且包括知识的增减，及时更新等进行处理，从一方面来讲，在处理知识库时，我们也应当使用协同机制，这样这两个数据库中，我们可以找到不一样的消息来源，这样再将结果提交给数据挖掘系统，我们就可以得到，自己想要的的数据，另一方面，我们应该将所知的所有知识库管理系统信息全都融入到知识库管理系统的总库里面去^[4]。

5.数据挖掘工具主要指的是在数据挖掘的过程中，我们应该使用到的一些模型或是推演方法，在这个推演过程中，我们应该对数据挖掘工具进行管理，了解到在整个数据挖掘的系统中存在着哪一些计算方法的工具，从而可以更加清楚的了解我们在挖掘数据时，应该利用哪些工具才能得到自己想要的的数据结果分析。

6.数据挖掘的预处理，主要是指在数据仓库的管理系统中，存在着子集还有母集，我们应该确定我们要从子集中提取数据，还是母集中提取数据，通过两者相关联系，对整个数据库中的所有数据进行处理，找到用户真正想要的样子，使用合理的挖掘工具就可以找到用户真正想要的的数据

7.在知识管理系统的模块中，有很多数据是比较枯燥无味的，因此，在数据管理系统中，我们应该及时摒弃那些不需要的，数据留下那些需要的数据，及时更新，甚至在更换数据时，我们应该采用数据挖掘的算法，重新计算好，我们应该采取的数据数量，因此，如果在知识库中存在一些无用的模式，那么，我们应该将其剔除

选用有用的模式来满足所有客户的基本要求

8.通过整个基本框架,我们可以知道,在整个发展模式的基本脉象,都是为了让客户得到有用的数据,首先,在查询过程中,我们应该运用索引的方法子集的方法,让客户得到有用的消息,其实我们在利用数据挖掘的过程中,应该使用正确的数据挖掘工具,数据挖掘工具系统应该即使展示所有的数据挖掘,在知识库方面,我们应该做到及时更新自己所有的数据,摒弃不需要的,不利于用户查询的模式^[9]。

六、基于数据仓库与数据挖掘技术的DSS

为了让数据仓库和数据挖掘更加有效的发展,企业的DSS应用,可以为其提供良好的决策,因此,我们创建数据仓库,就是为了让企业能够做出良好的科学决策,在数据仓库录入,大量的数据之后,我们该怎么样分析其DSS,首先介绍DSS应用主要包括哪一些内容?在数据库层面,我们应该将所有的数据全部整合到数据库当中,在数据库当中,我们应该在通过数据挖掘手段,通过多维数据模型集成转换,再把得到的结果录到数据仓库当中,通过数据字典进行刷新加载,将数据仓库的数据进一步挖掘,主要使用OLAP的方法和工具,工具层我们就可以直接采用DSS应用,主要是通过多维数据模型进行数据的集成和转换,批量处理数据的仓库,从而进行DSS的分析,为企业的DSS建立提供更有效的措施。

七、结束语

通过对数据仓库与数据挖掘教程的学习,我了解了数据仓库对相关数据进行分析的方法,以及用相关软件预测的步骤。并且懂得了企业未来的成功,很大程度上

取决于准确的数据挖掘能力,许多领域都需要对潜在的数据进行深层次的分析,困难主要有对数据的一些概念和方法不太熟悉,导致思想比较懵懂,会时不时出现差错,导致出现的结果与预期的不一致。但总体来说还是有很多收获的,通过这次学习,我巩固了所学的理论知识,进一步理解了相关的概念和方法。也明白了一些深刻的道理,即在遇到困难时不要放弃,要有持之以恒的精神,遇到不懂的问题时要及时请教老师和同学,要在实际动手操作时进一步完善自己的所学的知识,要善于思考,善于总结,这样才能有所学有所想,学有所得。

参考文献:

- [1]汪翀.基于Hadoop生态圈的区域基础教育大数据应用体系架构研究及实现[J].中国教育信息化(高教职教),2021(8):54-60.DOI:10.3969/j.issn.1673-8454.2021.08.011.
- [2]张利艳.基于企业如何利用信息技术进行统计工作转型的研究分析[J].科学与信息化,2021(17):5,7.
- [3]李思雨,乔鹏程.升级数据中台和建立智能数据仓库方式的绩效管理创新——以AL集团为例[J].航空财会,2021,3(5):65-71.DOI:10.19840/j.cnki.FA.2021.05.012.
- [4]胡晓曦.基于不动产GIS的办公用房用地管理一体化数据库模型应用研究[J].矿山测量,2020,48(6):69-76.DOI:10.3969/j.issn.1001-358X.2020.06.016.
- [5]何海,张小荣,赵沙沙,等.基于数据挖掘及网络药理学方法探讨中药复方抗上呼吸道感染的用药规律及作用机制[J].安徽农业科学,2022,50(3):178-185,212.DOI:10.3969/j.issn.0517-6611.2022.03.048.