

## 国内数据挖掘工具研究综述\*

## Review of Domestic Research on Data Mining

钱 峰

(南京大学商学院 南京 210093)

**摘 要** 在对国内数据挖掘工具的相关文献计量分析的基础上,综述了其三大研究主题:数据挖掘工具的分类、数据挖掘工具的介绍与开发、数据挖掘工具比较与评价,并归纳了数据挖掘工具研究的两个发展方向:专业领域数据挖掘工具的开发、数据挖掘工具的分类评价。

**关键词** 数据挖掘 数据挖掘工具 评价研究

**中图分类号** TP311.1

随着 20 世纪 80 年代“数据挖掘”这一新兴学科的出现,形形色色的数据挖掘商业软件工具也竞相面世。按照姚毓才等的观点,能够从巨量数据中获取有效的、新颖的、潜在有用的、最终可理解的模式的非平凡过程的软件工具可称为数据挖掘工具<sup>[1]</sup>。当前的数据挖掘工具有哪些?如何选用最合适的数据挖掘工具等问题是用户经常面对的问题。本文尝试在国内现有文献的基础上综述数据挖掘工具的研究主题,并归纳其发展方向。

## 1 数据挖掘工具研究的文献计量分析

数据挖掘已在多个领域中得到广泛应用,在 CNKI 与维普(VIP)中文科技期刊数据库中能检索到的最早的相关论文发表于 1997 年。软件是数据挖掘的必备工具,在 CNKI 与 VIP 中能检索到的最早的“数据挖掘工具”的相关论文发表于 1998 年。截至笔者撰写本文时,CNKI 中共收录相关论文 35 篇。这 35 篇论文按时间和学科的分布分别如表 1、表 2 所示。

表 1 “数据挖掘工具”相关论文的时间分布

年份	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
论文数(篇)	1	1	1	1	4	7	5	5	5	5

表 2 “数据挖掘工具”相关论文的学科分布

学者所属学科	计算机	图书情报学	遥感自动化	商学
论文数(篇)	25	4	4	2
所占比例	71.4%	11.4%	11.4%	5.7%

表 1 显示:自 1998 年起,CNKI 中“数据挖掘工具”的相关论文数量呈递增趋势。这表明数据挖掘工具逐渐成为学者们关注的对象。究其原因,主要在于数据挖掘在多个领域的广泛应用,而工具在数据挖掘的应用中必不可少,自然也越受关注。

表 2 显示:从事数据挖掘工具研究的学者主要分布在四个专业领域,即计算机、图书情报学、遥感自动化、商学。从论文

数量上看,计算机领域发表的相关论文数量最多(25 篇),占 71.4%,其次为图书情报学领域与遥感自动化领域,均占 11.4%,再次是商学领域,占 5.7%。究其原因,主要在于非计算机领域的学者通常只是将数据挖掘作为一种方法,从事研究工作时,会使用数据挖掘工具即可,而不会专门从事数据挖掘工具的研究。表 2 中的非计算机领域的学者所发表的相关论文主要是对数据挖掘工具的介绍或比较、评价,而未涉及数据挖掘工具的开发、设计、算法等内容。

## 2 数据挖掘工具研究的内容分析

通过分析近 10 年文献的主题内容,笔者将数据挖掘工具的研究主题归纳为三个方面:数据挖掘工具的分类、数据挖掘工具的介绍与开发、数据挖掘工具的比较与评价。分类研究是为了了解当前数据挖掘工具的共性与特性;开发则是因为当前的数据挖掘工具不能满足特定领域的特定需求;比较与评价研究是按照某种或某些性能筛选数据挖掘工具。

**2.1 数据挖掘工具的分类** 按照一般的分类方法,数据挖掘工具可以分为:特定领域的数据挖掘工具和通用的数据挖掘工具。特定领域的数据挖掘工具是针对某个特定领域的问题提供解决方案,其针对性比较强,往往采用一些特殊的算法,可以处理特殊的数据,实现特殊的目的。通用的数据挖掘工具不区分具体数据的含义,采用通用的挖掘算法,处理常见的数据类型,挖掘什么、用什么来挖掘都由用户根据自己的应用来选择。这两类数据挖掘工具的实例如表 3 所示。

按照软件所基于的平台划分,数据挖掘工具可分为:基于 DOS 的软件工具、基于 Windows 的软件工具、基于 Linux 的软件工具和基于 Solaris 的软件工具等。

此外,也有学者按照其他标准将数据挖掘工具划分为其他类别。例如,高英明等根据应用领域将数据挖掘工具分为 3 类:通用单任务类、通用多任务类、专用领域类;而根据所采用

基金项目:国家自然科学基金项目“人力资源管理与先进制造技术及其匹配对产品创新的影响”(编号:70740005)。

作者简介:钱 峰,男,1976 年,博士,研究方向为企业知识管理、数据挖掘与决策支持系统。

的技术将数据挖掘工具分为 6 类:基于规则和决策树的工具、基于神经网络的工具、数据可视化方法、模糊发现方法、统计方法和综合多方法<sup>[3]</sup>。

表 3 特定领域的与通用的数据挖掘工具实例

分类	数据挖掘工具	开发单位	功能
特定领域的数 据挖掘工具	AdvancedScout	IBM 公司	针对 NBA 的数据,帮助教练优化战术组合
	SKICAT 系统	加州理工学院喷气推进实验室与天文科学家合作开发	帮助天文学家发现遥远的类星体
	TASA	芬兰赫尔辛基大学计算机科学系	帮助预测网络通信中的警报
通用的数据 挖掘工具	Clementine	SPSS 公司	
	Enterprise Miner	SAS 公司	
	QUEST 系统	IBM 公司 Almaden 研究中心	通用型
	MineSet 系统	SGI 公司	
	DBMiner 系统	加拿大 Simon Fraser 大学	

注:表中部分数据来源于文献<sup>[3]</sup>

姚毓才等借鉴多维数据立方体表示的概念,提出了数据挖掘软件工具的多维立方体分类方案,即用数据挖掘工具的不同属性类作为多维立方体的维,从而建立一个多维的立方体模型,然后在此基础上进行有关知识的挖掘,该模型的三个维分别是:数据源(数据库、文本文件 Web 数据、多媒体数据等)、方法(统计方法、决策树、规则抽取、基于事例的推理、Bayes 网络、遗传算法、神经网络、模糊集、Rough 集等)和功能(预处理、关联分析、回归分析、分类、预测、聚类分析、异常分析、模型可视化等)。如图 1 所示。

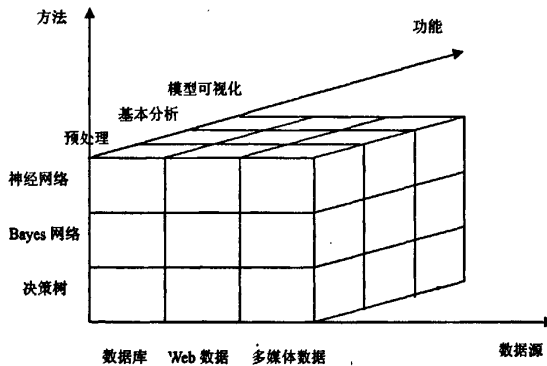


图 1 数据挖掘工具的三维立方体分类方案

2.2 数据挖掘工具的介绍与开发 通用数据挖掘工具已越来越不能满足各专业领域研究日益深入的需要,侧重研发的学者便自行开发新的数据挖掘工具,例如:

刘君强等<sup>[4]</sup>开发了智能型数据挖掘工具——SmartMiner,该工具是以在数据挖掘算法研究和专家系统研究中取得的成果为基础,提出了挖掘作业描述语言 MDL 和挖掘任务模型脚本语言,设计了挖掘向导、可视化向导和挖掘任务模型,集成了数据仓库管理功能,挖掘引擎具有智能性,体系结构开放并可扩展。袁红春等<sup>[5]</sup>将常用于关系数据库和事务数据库的面向属性的归纳和关联规则挖掘方法扩展应用于 GIS 数据库,开发了空间数据挖掘工具—GISMiner。何耀东等<sup>[6]</sup>开发了 DM Tools,实现了基于数据库的知识发现(KDD)的主要过程:数据

可视分析、数据预处理、数据库的知识发现、数据挖掘、模型解释及模型评估等,该工具可从各行业的历史业务数据库中挖掘出隐含的有价值的知识,用于决策支持。刘伟<sup>[7]</sup>、田小霞<sup>[8]</sup>等学者设计了基于 CORBA 的数据挖掘工具。而侧重应用的学者则主要介绍企业或学术机构研发的数据挖掘工具的功能应用,例如:邓英等介绍了两类 Web 数据挖掘工具<sup>[9]</sup>:文本信息挖掘工具,如 IBM 公司推出的 Intelligent Miner for Text;用户访问模式挖掘工具,如用户导航行为挖掘工具 WUM (web Utilization Miner, <http://wum.wiwi.hu-berlin.de> 上可免费获得),以及 Stephen Turner 博士编制的免费个人软件 Analog (<http://www.statlab.com.ac.uk> 上可免费获得)。汤海鹏等介绍了专业航空遥感图像处理系统 Erdas 和 DB Miner 的特点<sup>[10]</sup>。Erdas 是由 ERDAS 等公司开发的专用于遥感图像处理的一个系统,提供的主要功能有:非监督分类、监督分类、知识分类、知识引擎等。DBMiner 是加拿大 Simon Fraser 大学数据库科研究实验室下的数据挖掘科研小组开发的一个多任务数据挖掘系统。王颖等介绍了可视化数据挖掘工具——RefViz 的基本功能:数据导入与导出、数据分析、顾问、词表工具等,并以肺炎支原体为例介绍了该工具的用法<sup>[11]</sup>。

2.3 数据挖掘工具的比较与评价 数据挖掘工具的比较研究主要包括基本属性的比较与功能的比较,数据挖掘工具的评价研究主要为评价指标体系研究。比较与评价的目标相同,都是为了筛选数据挖掘工具。

表 4 数据挖掘工具的比较研究实例

作者	比较对象	比较属性	主要结论
蒋秋林	Intelligent Miner, DB Miner 2.0, SAS, Darwin	产品提供商、用户界面、拥有成本、商业竞争能力、建模技术与算法、模型发布	Intelligent Miner 简单易用,能进行大数据量的挖掘,功能一般,没有数据探索功能,与其他软件接口差,难于发布,结果虽美观但不理解;DB Miner 2.0 算法简单,价格便宜,系统功能较全面,并提供了开放式体系结构;SAS 系列产品功能强大,但难以掌握;Darwin 功能较弱,使用不方便,没有数据探索功能,市场份额小
张雪英	Clementine - SPSS, Intelligent Miner, Darwin, Teradata Miner, SQL Server 2000, SAS, DB Miner 2.0	产品提供商、方法论、商业理解、支持平台数据展现、用户界面、提供商业竞争能力、拥有成本、数据准备、建模技术与算法、评估、模型发布	SAS 完全以统计理论为基础,功能强大,有完备的数据探索功能,要求是高统计分析专业人员,价格也极其昂贵;IBM 的 Intelligent Miner 简单易用,能处理大数据量的挖掘,功能一般,没有数据探索功能,与其他软件接口差;Oracle 功能较弱,使用不方便,没有数据探索功能,市场份额也小;SPSS 功能强大,有完备的数据探索功能,也较易掌握,性价比较高,有能力处理大数据量,而且 SPSS 具有方便的发布和集成功能;Teradata 新推出自己的挖掘工具,在某些方面还不太成熟
马彦中	SPSS, STATISTICA, Tiberius	回归分析	SPSS 支持脚本功能,鼓励人工参与,支持多种类型数据,提供丰富的图形表示功能;STATISTICA 易于学习和使用,功能单一;Tiberius 功能强大,支持几乎所有数据挖掘技术,具有批处理功能

2.3.1 数据挖掘工具的比较。具有代表性的数据挖掘工

具的比较研究包括郭秋萍<sup>[12]</sup>、张雪英<sup>[13]</sup>、马准中<sup>[14]</sup>等人的研究。他们从不同的方面对国内外常用的数据挖掘工具进行了比较,比较的属性与结论如表 4 所示。

2.3.2 数据挖掘工具的评价。当前数据挖掘工具的评价研究主要为综合评价,不同的学者从不同的角度设计了不同的分类体系,但评价对象均是所有数据挖掘工具,具体如表 5 所示。

表 5 数据挖掘工具的评价体系实例

作者	评价指标
朱莹	商业理解,数据理解,建模,模型评估,结果发布
高英明 刘克勤	系统功能及工具的完备性,系统体系结构的可扩展性,多种数据源的存取能力,底层数据模型支持能力,数据量的最大处理能力,提供给用户的表现能力
郭秋萍	数据挖掘工具的功能及其完备性,数据挖掘工具的可扩展性,数据挖掘工具的数据准备和处理能力,解决复杂问题的能力,数据挖掘工具提供给用户的表现力,数据挖掘工具的易操作性,系统的开放性
周德吉	数据存储,数据处理,模型算法,自动建模,可视化,其他
邓英	可产生的模式种类的多少,解决复杂问题的能力,易操作性,数据存取能力,与其他产品的接口
李建 刘红星	数据准备,数据访问,算法与建模,模型评价和解释,用户界面

事实上,随着各种特定领域的数据挖掘工具的出现,通用的、综合型的评价体系将不再是研究的重点,学者们将越来越关注对特定领域的数据挖掘工具的评价研究。

综上所述,数据挖掘工具的分类研究、数据挖掘工具的介绍与开发研究、数据挖掘工具的比较与评价研究概括了国内数据挖掘工具的研究现状。

### 3 数据挖掘工具研究的方向

根据学者与数据挖掘工具之间的关系,可将学者群体分为两类:开发者与使用者。从开发者的角度而言,开发专业领域的数据挖掘工具将成为一个重要的研究方向,因为数据挖掘日益广泛地应用在过程中,必将有越来越专业化的需求,如 Web 挖掘中对于音频、视频等多媒体数据的挖掘。从使用者的角度而言,数据挖掘工具的分类评价也将成为一个重要的研究方向,越来越多的专业领域的数据挖掘工具将使得数据挖掘工具的综合评价适用性降低,分类评价将成为关注的焦点。

3.1 专业领域数据挖掘工具的开发 数据挖掘的研究热点包括:Web 挖掘(Web 内容挖掘、Web 结构挖掘、Web 日志挖掘)、生物信息挖掘、文本数据挖掘等<sup>[15]</sup>。在这些领域,数据挖掘工具也必将广泛使用,但通用数据挖掘工具通常难以胜任这些领域的数据挖掘。

数据库有规范的结构,其中数据有统一的格式,而且高度结构化,而 Web 上的数据结构性差,多为半结构化或非结构化,由于半结构化和非结构化的信息不能清楚地用数据模型来表示,因此在 Web 挖掘中开发更适用 Web 数据(如 Web 多媒体数据、链接数据等)的数据挖掘工具便成为必然。在遥感领域,通过卫星扫描地球,每天都能获得大量的关于地表的遥感图像,要从大量的数据中判读出每一个图片所潜藏的信息,必然用到空间数据挖掘技术。在空间数据挖掘过程中,通用数据挖掘工具仍显不足,针对空间数据特征,开发专业的空间数据挖掘工

具也将成为必然。其他领域亦然。

3.2 数据挖掘工具的分类评价 一方面,数据挖掘工具在多个领域体现出越来越重要的作用;另一方面,各专业领域的数据挖掘工具日益丰富。因此,如何选择合适的数据挖掘工具必将成为用户不得不面对的问题。针对各专业领域的不同需求,将不同领域的数据挖掘工具按照统一的指标体系评价是不可行的,因为这样的评价结果并不能为用户筛选出合适的工具。而数据挖掘工具的分类评价则可以解决这一问题。分类评价首先需要按照一定的标准将数据挖掘工具分类,然后再按照预先设定的评价指标体系对各类工具进行评价。分类评价的结果的针对性更强,可有效地为特定领域的用户选择特定的工具。

### 4 结束语

对数据挖掘工具的研究仅有十余年的历史,作为数据挖掘领域的一个研究分支,尽管其作用不言而喻,却并未引起从事数据挖掘研究的学者的充分注意。笔者以国内数据挖掘工具的相关文献为基础,综述其发展现状,并归纳其发展方向,以期能引起更多学者对数据挖掘工具的关注。

#### 参考文献

- 姚毓才,王本年.数据挖掘工具的分类与挖掘[J].计算机技术与发展,2006,16(8):6-9
- 数据挖掘工具简介[EB].2008-2-23. <http://forest-1980.blog.163.com/blog/static/9740216200710211730707/>
- 高英明,刘克勤.基于数据仓库的数据挖掘工具[J].现代电力,2002,19(2):71-75
- 刘君强,王 勤,孙晓莹.智能型数据挖掘工具的设计与实现[J].计算机工程与应用,2003(17):195-197
- 袁红春,熊范纶,杭小树等.一个适用于地理信息系统的数据挖掘工具—GISMiner[J].中国科学技术大学学报,2002,32(2):217-224
- 何耀东,常桂然,徐 茜等.数据挖掘工具 DM Tools 的设计与实现[J].中国图像图形学报,1999,4(11):936-940
- 刘 伟,刘同明,曹奇英等.基于 CORBA 的数据挖掘工具 KDD-DC[J].计算机应用研究,2000(7):34-35
- 田小霞,刘晓霞,钟少丹.基于 CORBA 的 Web 数据挖掘工具的设计及应用[J].现代计算机,2003(2):26-28
- 邓 英,李 明.Web 数据挖掘技术及工具研究[J].计算机工程与应用,2001(20):92-94
- 汤海鹏,毛克彪,覃志豪等.空间数据挖掘工具浅谈[J].测绘与空间地理信息,2005,28(3):4-6
- 王 颖,戎文慧.可视化文本分析和数据挖掘工具 RefViz[J].中华医学图书情报杂志,2006,15(6):61-64
- 郭秋萍.企业数据挖掘工具的分析与选择[J].航空档案,2004(10):80-82
- 张雪英.国外先进数据挖掘工具的比较分析[J].计算机工程,2003,29(16):1-3
- 马准中.对三种数据挖掘工具的比较[J].实验科学与技术,2005(1):37-39

(责编:王平军)

# 国内数据挖掘工具研究综述

作者: [钱峰, Qian Feng](#)  
作者单位: [南京大学商学院, 南京, 210093](#)  
刊名: [情报杂志](#) **PKU** **CSSCI**  
英文刊名: [JOURNAL OF INFORMATION](#)  
年, 卷(期): 2008, 27(10)  
被引用次数: 0次

## 参考文献(14条)

1. [姚毓才, 王本年](#) 数据挖掘工具的分类与挖掘[期刊论文]-[计算机技术与发展](#) 2006(08)
2. [数据挖掘工具简介](#) 2008
3. [高英明, 刘克勤](#) 基于数据仓库的数据挖掘工具[期刊论文]-[现代电力](#) 2002(02)
4. [刘君强, 王勋, 孙晓莹](#) 智能型数据挖掘工具的设计与实现[期刊论文]-[计算机工程与应用](#) 2003(17)
5. [袁红春, 熊范纶, 杭小树](#) 一个适用于地理信息系统的数据挖掘工具-GISMiner[期刊论文]-[中国科学技术大学学报](#) 2002(02)
6. [何耀东, 常桂然, 徐茜](#) 数据挖掘工具DM Tools的设计与实现[期刊论文]-[中国图象图形学报](#) 1999(11)
7. [刘伟, 刘同明, 曹奇英](#) 基于CORBA的数据挖掘工具KDD-DC[期刊论文]-[计算机应用研究](#) 2000(07)
8. [田小霞, 刘晓霞, 钟少丹](#) 基于CORBA的Web数据挖掘工具的设计及应用[期刊论文]-[现代计算机](#) 2003(02)
9. [邓英, 李明](#) Web数据挖掘技术及工具研究[期刊论文]-[计算机工程与应用](#) 2001(20)
10. [汤海鹏, 毛克彪, 覃志豪](#) 空间数据挖掘工具浅谈[期刊论文]-[测绘与空间地理信息](#) 2005(03)
11. [王颖, 戎文慧](#) 可视化文本分析和数据挖掘工具RefViz[期刊论文]-[中华医学图书情报杂志](#) 2006(06)
12. [郭秋萍](#) 企业数据挖掘工具的分析与选择 2004(10)
13. [张雪英](#) 国外先进数据挖掘工具的比较分析[期刊论文]-[计算机工程](#) 2003(16)
14. [马准中](#) 对三种数据挖掘工具的比较[期刊论文]-[实验科学与技术](#) 2005(01)

## 相似文献(10条)

### 1. 学位论文 [周宇葵](#) 国外数据挖掘与知识发现工具的评估研究 2004

背景:经过十多年发展,国外数据挖掘与知识发现工具已大量涌现并开始大幅度由研究走向应用。国内数据挖掘工具研究虽处在起步阶段,但应用需求也很大。如何借鉴国外先进经验与成果进行研究,如何选择数据挖掘工具以促进应用,这是国内数据挖掘领域发展所面临的一个紧迫问题。目的:本研究旨在通过对国外主流数据挖掘工具的评估了解国外数据挖掘工具的发展现状,分析数据挖掘工具发展过程中存在的问题,对数据挖掘工具的发展提出若干建议,并试图阐述数据挖掘工具的发展给情报工作带来的影响。方法:本文以国外主流数据挖掘工具为研究对象,根据国内外研究成果以及国外数据挖掘与知识发现软件的发展状况,设计了一套由7个部分共119项组成的评估表,并用该表对国外主流数据挖掘与知识发现软件进行了比较与评估。结论:就总体特征而言,迄今为止,通用数据挖掘软件的使用占绝对优势;数据挖掘与知识发现软件的功能不再停留在“挖掘”这个单一功能的实现,而已延伸到数据挖掘与知识发现的广义过程,即包括数据的预处理、数据挖掘、模型评估与可视化;数据挖掘软件使用的专业性较强;数据挖掘对象较单一;数据挖掘可视化取得了一定的进展,在单纯的模型可视化基础上扩充了数据可视化与数据挖掘过程可视化;数据挖掘还未真正融入企业应用与决策。就具体方面来说,各软件在能访问的数据源类型,能完成的数据挖掘任务,数据预处理,参数设置的灵活性,提供的算法的多样性与层次性,数据、过程以及模型的可视化程度等方面各有特色。本文在比较评估基础上对数据挖掘软件的发展提出若干建议,包括完善专用数据挖掘工具功能;完善数据挖掘标准化的内容,促进已有数据挖掘标准的应用;扩充数据挖掘对象的范围,实现对对象数据库、空间数据库等非结构化数据的挖掘;实现数据挖掘与企业内部其它应用软件的无缝集成;各数据挖掘工具应相互借鉴,取长补短;加强数据挖掘与数据库、数据仓库等的一体化研究;提供对移动数据的挖掘;实现互联网上的数据挖掘等。

### 2. 学位论文 [贺伟淞](#) 可视化数据挖掘工具的设计与实现 2005

论文首先介绍了可视化数据挖掘工具的研究背景、现状、意义、来源、目标和论文工作,可视化数据挖掘领域内国内外已有的文献综述;接下来,论文介绍了数据挖掘和可视化数据挖掘的基本概念、技术与方法;然后介绍了可视化数据挖掘系统的原型结构,在此基础上设计出可视化数据挖掘系统(VDMTOOLS)的架构,进而给出了系统的JAVA实现,并对系统进行了仿真测试和分析。最后,总结了已经完成的工作,对仍然需要解决的问题提出解决方案,提出今后的可行的研发设想。

### 3. 学位论文 [李忍](#) 一个体育比赛技战术分析数据挖掘工具的设计与实现 2008

随着社会的发展和人类的进步,体育竞技现在已经成为一个国家和民族文明进步的一个重要标志。为了体育竞技水平的提高,越来越多的国家不断将新的科技应用到体育竞技之中。数据挖掘就是从大规模的数据中抽取出自非平凡的、隐含的、未知的、有潜在使用价值的信息的技术。因此,将数据挖掘技术应用于体育比赛,对比赛数据进行分析,将对体育竞技水平的提高具有不可估量的作用。

本文在学习实验室原有乒乓球技战术分析系统及排球技战术分析系统,并对数据挖掘技术深入研究的基础上,提取出通用的数据挖掘算法,生成开发类库中间件,作为体育比赛技战术分析工具,可以方便地加入到技战术分析系统当中,并将该工具应用到乒乓球技战术分析及排球技战术分析中去。本文的主要内容如下:

首先,是研究背景和现状,重点介绍了国内外数据挖掘技术在体育比赛中的应用情况。以及对实验室原有的乒乓球技战术分析系统及排球技战术分析系统进行了简单说明。

其次,作为本文研究的理论基础,介绍了数据挖掘的概念、发展过程、研究现状,以及对中间件技术的作了简单介绍。

另外,在系统分析及系统设计的基础上,对关联规则分析、聚类分析和基于马尔科夫过程分析三种挖掘算法进行了改进,并抽象出通用的开发类库,使之可以作为工具方便地加入到球类比赛技术分析中去。

然后,结合乒乓球技术分析系统及排球技术分析系统,分别将挖掘算法引入到技术分析系统中去。结合所采集的比赛数据,进行技术分析,并将挖掘结果进行清晰展示,可以方便进行比赛指导。

最后,结合乒乓球技术分析系统及排球技术分析系统,对实现的开发类库进行了测试。

#### 4. 期刊论文 [张雪英](#) [国外先进数据挖掘工具的比较分析](#) -[计算机工程](#)2003, 29(16)

近年来,国外陆续推出了一些先进的数据挖掘工具,国内也在不断地引入这些数据挖掘工具。随着数据挖掘工具的不断涌现,如何选择适合企业自身特定需要的数据挖掘工具,已成为企业引入数据挖掘技术的一大难题。文章在简要概述数据挖掘技术背景的基础上,从企业应用的角度,全面详细地比较分析了当前国外先进的数据挖掘工具。

#### 5. 学位论文 [王冬磊](#) [数据挖掘技术在化工工艺优化中的应用研究](#) 2005

随着数据存储量的飞速增长,传统的分析技术已经不能满足需求,于是将数据挖掘技术应用到工程领域,通过数据挖掘工具对实验数据进行关联规则挖掘,从而对相应的工艺参数进行优化用以实现工程信息的智能化就变得更加必要。本文将数据挖掘技术应用到流化床电极反应器应用研究中,并取得了一定成效,实验结果表明通过数据挖掘工具对实验数据进行关联规则挖掘,进而对相应的工艺参数进行优化,是一个很好的途径,将明显的提高其经济效率。

本文的主要工作成果如下:

(1)介绍了数据挖掘和数据挖掘的相关技术。概述了数据挖掘的过程,重点的描述了几个常用数据挖掘功能与数据挖掘方法,对如何评估数据挖掘的结果也做了些相应的介绍。

(2)介绍了数据挖掘技术在化学化工领域的应用研究现状,并对其应用前景进行了分析。

(3)对流化床电极电解实验的背景和相关的理论进行了介绍,并进行了相关的流化床电极电解实验,得出实验数据,以进行分析。

(4)介绍了过去几年里国外的一些公司和研究所开发的较好的数据挖掘工具,分析和对比了其中相对比较重要的挖掘工具,并在此基础上讨论了如何根据特定公司的商业策略选择一种最适合的数据挖掘工具。

(5)最后,应用一个基于excel的数据挖掘工具idata对流化床电极实验的数据进行关联规则挖掘。并得出有兴趣的规则。

本文的组织如下:第一章介绍数据挖掘和数据挖掘的相关技术。第二章介绍了数据挖掘技术在化学化工领域的应用研究现状。第三章介绍了流化床电极电解实验的情况。第四章分析了几个典型的数据挖掘工具。第五章对实验数据进行挖掘,得出有兴趣的规则。第六章总结全文和对未来工作展望。

#### 6. 期刊论文 [周宇葵](#),[刘莉](#).[Zhou Yukui.Liu Li](#) [数据挖掘工具能力评估研究](#) -[中国中医药现代远程教育](#)2005, 3(12)

数据挖掘是从大量的、不完全的、有噪声的、模糊的、随机的数据中提取隐含在其中的、人们事先不知道的但又是潜在有用的信息和知识的过程。数据挖掘工具是实现数据挖掘并应用数据挖掘的直接产物,要做数据挖掘必须通过数据挖掘工具。因此,数据挖掘工具的优劣直接关系到用户对企业内部数据应用需求的实现程度与范围。

#### 7. 学位论文 [万家华](#) [基于SEMMA的数据挖掘工具设计及其关键技术研究](#) 2001

通用的数据挖掘工具必须支持从数据准备、数据变换到模型评估这一整个数据挖掘过程,而不仅仅是一些数据挖掘算法的集成。为了解决这一问题,该文研究了数据挖掘方法论SEMMA以及设计通用数据挖掘工具的一些关键技术。该文首先在比较数据挖掘工具和一般数据库工具,分析主流数据挖掘工具的基础上,给出了数据挖掘工具的基本要求,接着讨论了数据挖掘方法论SEMMA,它将整个数据挖掘过程分为5步,即Sample(采样),Explore(探索),Modify(修改),Model(建模)和Assessment(评估),反应了数据挖掘的本质。然后基于SEMMA方法论,重点讨论了设计通用数据挖掘工具的关键技术,包括数据挖掘过程建模、可视化、模型评估和模型应用。

#### 8. 学位论文 [逢焕利](#) [数据挖掘工具软件研究与实现](#) 2002

计算机技术的飞速发展把人们带入了信息社会,并且致使信息的拥有量及其增长速度极为惊人。如何从巨量的信息中提取人们感兴趣的知识成为亟待解决的问题。由此数据挖掘就成为数据库领域最富活力的研究方向,并在许多领域得到重要应用。数据挖掘的研究至今已有一十年左右的历程,这期间致力于这个课题研究的学者们提出了许多挖掘方法,并做了许多实验,也开发了一些工具,但绝大多数研究始终未能走出实验室,投入使用。还缺少适合中国国情的有自主知识产权的数据挖掘工具软件。为了有计划地、系统地研究和开发出实用型的数据挖掘工具软件,我们在吉林省教委立项研究和开发数据挖掘工具软件。该文讨论了四种数据挖掘算法—关联规则、决策树、粗糙集和遗传算法—的基本理论和算法,并对经典关联规则算法和遗传算法进行了改进。在此基础上,设计开发了数据挖掘工具软件,从而为构造有效的数据挖掘平台打下基础。

#### 9. 期刊论文 [马准中](#).[Ma Zhunzhong](#) [对三种数据挖掘工具的比较](#) -[实验科学与技术](#)2005, 3(1)

在简要概述数据挖掘技术背景的基础上,从回归分析技术应用的角度,全面详细地比较分析了当前国外先进的三种数据挖掘工具。

#### 10. 期刊论文 [贺红](#),[蔡晓莲](#) [数据挖掘及其工具的合理选择](#) -[福建电脑](#)2009, 25(1)

本文给出了基于数据库的知识发现(KDD)的核心—数据挖掘,通过对数据挖掘的实现过程和种类的分析,给出了数据挖掘工具的合理选择标准,为充分利用数据挖掘工具提供了参考。

本文链接: [http://d.g.wanfangdata.com.cn/Periodical\\_qbzz200810003.aspx](http://d.g.wanfangdata.com.cn/Periodical_qbzz200810003.aspx)

授权使用: 中国科技大学(wfzykjdx), 授权号: f6674466-9a52-4052-8d2a-9e72011e4354

下载时间: 2011年1月21日