

<<关联规则技术研究>>

图书基本信息

书名：<<关联规则技术研究>>

13位ISBN编号：9787308099264

10位ISBN编号：7308099261

出版时间：2012-6

出版单位：浙江大学出版社

作者：沈斌

页数：169

字数：170000

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<关联规则技术研究>>

内容概要

《关联规则技术研究》是作者沈斌在关联规则挖掘领域多年研究成果的系统总结。在相关综述的基础上，重点探讨了以下几个方面的工作：提出了一种新的相关性兴趣度量All-item-confidence，研究了该度量所具有的性质，然后阐述了该度量与All-set-confidence度量之间的关系，以及该度量的适用范围；针对普通关联规则在解决前后项集对称型应用问题上的不足，提出了项项正相关关联规则挖掘问题；针对普通关联规则在解决前后项集非对称型应用问题上的不足，进一步提出了项项且项集正相关关联规则挖掘问题；对动态关联规则技术进行了相应的研究，提出了动态关联规则新定义，进一步阐述了两种动态关联规则挖掘新算法；提出了挖掘带使用信息的动态关联规则(DAR-C)新问题，给出了DAR-C规则的候选有效时段的表示方法，并对DAR-C规则进行了定义等。

本书一方面是数据挖掘相关的科技工作者学习、研究、应用、推广关联规则技术的重要参考资料；另一方面也可以作为相关领域研究生、本科生的教材和学习资料。希望本书能够促进广大科技工作者对关联规则的认识、应用和创新。

<<关联规则技术研究>>

作者简介

沈斌，工学博士，浙江大学宁波理工学院副教授，浙江大学硕士生导师，澳大利亚悉尼科技大学量子计算与智能系统研究中心访问学者。

近年来，主持或参与多项国家和省部级科研项目，发表高水平学术论文20余篇，出版专著1部，主编科普读物1部。

2011年被列入浙江省“新世纪151人才工程”第三层次培养人员。

主要研究方向：数据挖掘、物联网和智慧城市建设、复杂网络与复杂系统。

<<关联规则技术研究>>

书籍目录

第1章 概论

- 1.1 引言
- 1.2 关联规则技术基础知识
 - 1.2.1 基本概念
 - 1.2.2 挖掘方法
- 1.3 关联规则技术研究分类
 - 1.3.1 频繁模式、最大频繁模式和闭合频繁模式挖掘
 - 1.3.2 多种扩展形式的关联规则挖掘研究
 - 1.3.3 关联规则挖掘后处理
- 1.4 问题的提出

第2章 基于相关兴趣度的关联规则挖掘

- 2.1 引言
- 2.2 一种新的相关兴趣度度量
 - 2.2.1 已有的关联和相关兴趣度度量
 - 2.2.2 All-item-confidence相关兴趣度度量
 - 2.2.3 All-item-confidence与All-set-confidence之间的关系
 - 2.2.4 All-item-confidence与卡方检验之间的关系
- 2.3 购物篮中关联规则的应用类型分析
- 2.4 基于All-item-confidence的项项正相关关联规则挖掘
 - 2.4.1 项项正相关关联规则挖掘问题的提出
 - 2.4.2 兴趣度度量的选取
 - 2.4.3 基于All-item-confidence度量的项项正相关关联规则挖掘
- 2.5 挖掘算法ItemCoMine_AP和ItemCoMine_CT
 - 2.5.1 ItemCoMine_AP算法
 - 2.5.2 ItemCoMine_CT算法
 - 2.5.3 实验测评和比较分析
- 2.6 基于All-item-confidence和项集相关性度量的项项且项集正相关关联规则挖掘
 - 2.6.1 项项且项集正相关关联规则挖掘问题的提出
 - 2.6.2 项集相关性度量
 - 2.6.3 项项且项集正相关关联规则定义和举例
- 2.7 挖掘算法I&ISCoMine_AP和I&ISCoMine_CT
 - 2.7.1 I&ISCoMine_AP算法
 - 2.7.2 I&ISCoMine_CT算法
 - 2.7.3 实验测评和比较分析
- 2.8 本章小结

第3章 动态关联规则挖掘

- 3.1 引言
- 3.2 问题描述及其分析
 - 3.2.1 动态关联规则原定义
 - 3.2.2 原定义的不足之处
- 3.3 动态关联规则新定义及其挖掘算法
 - 3.3.1 动态关联规则新定义
 - 3.3.2 动态关联规则挖掘算法
 - 3.3.3 性能评测
- 3.4 带使用信息动态关联规则挖掘问题的提出

<<关联规则技术研究>>

3.5 问题定义

3.5.1 候选有效时段表示

3.5.2 带使用信息的动态关联规则

3.6 带使用信息的动态关联规则挖掘算法

3.6.1 挖掘框架

3.6.2 ITS2算法

3.6.3 EFP-GrOWth2算法

3.6.4 使用信息生成

3.6.5 性能评测

3.6.6 应用实例

3.7 本章小结

第4章 加权模糊层次关联规则挖掘

4.1 引言

4.2 模糊层次型关联规则

4.3 布尔型数据库中的加权模糊层次型关联规则挖掘

4.3.1 加权的原因

4.3.2 叶子结点项权值的确定

4.3.3 加权模糊层次型关联规则(WGF—AR)模型

4.4 WGF-AR规则挖掘算法

4.4.1 性质

4.4.2 W-Apriori算法

4.5 性能测评

4.5.1 实验一：算法性能测试

4.5.2 实验二：可伸缩性实验

4.6 本章小结

第5章 基于模糊分类结构的交易数据库关联规则聚类

5.1 引言

5.2 模糊分类结构的合并

5.2.1 模糊分类结构描述

5.2.2 多个有向无环图的合并

5.2.3 合并后的模糊分类结构描述

5.3 带语义差别信息的模糊分类结构

5.4 基于模糊分类结构的距离度量

5.4.1 项间距离

5.4.2 项集距离

5.4.3 关联规则距离

5.5 规则聚类算法的选择和应用

5.6 实验分析与讨论

5.6.1 实验一：规则距离计算实验

5.6.2 实验二：规则聚类可视化计算

5.7 本章小结

第6章 使用erot6g6软件的基于Ontology的关联规则检索

6.1 引言

6.2 相关概念简介

6.2.1 语义本体

6.2.2 语义网

6.2.3 关联规则及其度量

<<关联规则技术研究>>

6.3 基于Ontology的智能规则检索系统体系结构

6.3.1 体系结构

6.3.2 检索方式

6.4 基于语义web的关联规则检索核心技术

6.4.1 规则检索Ontology

6.4.2 商品项目实例和Rules实例标注

6.4.3 查询解析

6.5 本章小结

第7章 关联规则技术进展及趋势展望

7.1 最新技术进展

7.1.1 关联规则隐藏(Association Rules Hiding)

7.1.2 比对模式(Contrast Patter / Emerging Patter)

7.1.3 图模式(Graph Patter)

7.1.4 可行动关联规则(Actionable Association Rule)、领域驱动关联规则

7.1.5 关联规则、模式应用研究

7.2 值得关注的方向

附录 关联规则研究资料汇总

参考文献

<<关联规则技术研究>>

章节摘录

两个阶段关联规则挖掘过程主要包含两个阶段：第一阶段必须先从资料集合中找出所有的高频项目组(Frequent Itemsets)，第二阶段再由这些高频项目组中产生关联规则(Association Rules)。

关联规则挖掘的第一阶段必须从原始资料集合中，找出所有高频项目组(Large Itemsets)。

高频的意思是指某一项目组出现的频率相对于所有记录而言，必须达到某一水平。

一项目组出现的频率称为支持度(Support)，以一个包含A与B两个项目的2-itemset为例，我们可以经由公式(1)求得包含{A,B}项目组的支持度，若支持度大于等于所设定的最小支持度(Minimum Support)门槛值时，则{A,B}称为高频项目组。

一个满足最小支持度的k-itemset，则称为高频k-项目组(Frequent k-itemset)，一般表示为Large k或Frequent k。

算法并从Large k的项目组中再产生Large k+1，直到无法再找到更长的高频项目组为止。

关联规则挖掘的第二阶段是要产生关联规则(Association Rules)。

从高频项目组产生关联规则，是利用前一步骤的高频k-项目组来产生规则，在最小信赖度(Minimum Confidence)的条件门槛下，若一规则所求得的信赖度满足最小信赖度，称此规则为关联规则。

例如：经由高频k-项目组{A,B}所产生的规则AB，其信赖度可经由公式(2)求得，若信赖度大于等于最小信赖度，则称AB为关联规则。

案例分析就沃尔玛案例而言，使用关联规则挖掘技术，对交易资料库中的纪录进行资料挖掘，首先要设定最小支持度与最小信赖度两个门槛值，在此假设最小支持度min_support=5%且最小信赖度min_confidence=70%。

因此符合此该超市需求的关联规则将必须同时满足以上两个条件。

若经过挖掘过程所找到的关联规则「尿布，啤酒」，满足下列条件，将可接受「尿布，啤酒」的关联规则。

用公式可以描述Support(尿布，啤酒) $\geq$ 5%且Confidence(尿布，啤酒) $\geq$ 70%。

其中，Support(尿布，啤酒) $\geq$ 5%于此应用范例中的意义为：在所有的交易纪录资料中，至少有5%的交易呈现尿布与啤酒这两项商品被同时购买的交易行为。

Confidence(尿布，啤酒) $\geq$ 70%于此应用范例中的意义为：在所有包含尿布的交易纪录资料中，至少有70%的交易会同时购买啤酒。

因此，今后若有某消费者出现购买尿布的行为，超市将可推荐该消费者同时购买啤酒。

这个商品推荐的行为则是根据「尿布，啤酒」关联规则，因为就该超市过去的交易纪录而言，支持了“大部份购买尿布的交易，会同时购买啤酒”的消费行为。

从上面的介绍还可以看出，关联规则挖掘通常比较适用与记录中的指标取离散值的情况。

如果原始数据库中的指标值是取连续的数据，则在关联规则挖掘之前应该进行适当的数据离散化（实际上就是将某个区间的值对应于某个值），数据的离散化是数据挖掘前的重要环节，离散化的过程是否合理将直接影响关联规则的挖掘结果。

<<关联规则技术研究>>

后记

关联规则发掘技术在国内外的应用就目前而言，关联规则挖掘技术已经被广泛应用在西方金融行业企业中，它可以成功预测银行客户需求。

一旦获得了这些信息，银行就可以改善自身营销。

现在银行天天都在开发新的沟通客户的方法。

各银行在自己的ATM机上就捆绑了顾客可能感兴趣的行产品信息，供使用本行ATM机的用户了解。如果数据库中显示，某个高信用限额的客户更换了地址，这个客户很有可能新近购买了一栋更大的住宅，因此会有可能需要更高信用限额，更高端的新信用卡，或者需要一个住房改善贷款，这些产品都可以通过信用卡账单邮寄给客户。

当客户打电话咨询的时候，数据库可以有力地帮助电话销售代表。

销售代表的电脑屏幕上可以显示出客户的特点，同时也可以显示出顾客会对什么产品感兴趣。

同时，一些知名的电子商务站点也从强大的关联规则挖掘中的受益。

这些电子购物网站使用关联规则中规则进行挖掘，然后设置用户有意要一起购买的捆绑包。

也有一些购物网站使用它们设置相应的交叉销售，也就是购买某种商品的顾客会看到相关的另外一种商品的广告。

但是目前在我国，“数据海量，信息缺乏”是商业银行在数据大集中之后普遍所面对的尴尬。

目前金融业实施的大多数数据库只能实现数据的录入、查询、统计等较低层次的功能，却无法发现数据中存在的各种有用的信息，譬如对这些数据进行分析，发现其数据模式及特征，然后可能发现某个客户、消费群体或组织的金融和商业兴趣，并可观察金融市场的变化趋势。

可以说，关联规则挖掘的技术在我国的研究与应用并不是很广泛深入。

近年来关联规则发掘技术的一些研究由于许多应用问题往往比超市购买问题更复杂，大量研究从不同的角度对关联规则做了扩展，将更多的因素集成到关联规则挖掘方法之中，以此丰富关联规则的应用领域，拓宽支持管理决策的范围。

如考虑属性之间的类别层次关系，时态关系，多表挖掘等。

近年来围绕关联规则的研究主要集中于两个方面，即扩展经典关联规则能够解决问题的范围，改善经典关联规则挖掘算法效率和规则兴趣性。

<<关联规则技术研究>>

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>