

学校编码: 10384

分类号 \_\_\_\_\_ 密级 \_\_\_\_\_

学号: 200310040

UDC \_\_\_\_\_

廈門大學

硕士学位论文

基于模糊集的关联规则及其应用

Fuzzy Association Rule and Its Application

郑晶晶

指导教师姓名: 朱建平 教授

专业名称: 数量经济学

论文提交日期: 2006年4月

论文答辩时间: 2006年6月

学位授予日期: 2006年 月

答辩委员会主席: \_\_\_\_\_

评阅人: \_\_\_\_\_

2006年4月

# 厦门大学学位论文原创性声明

兹提交的学位论文，是本人在导师指导下独立完成的研究成果。  
本人在论文写作中参考的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文而产生的权利和责任。

声明人（签名）：

年 月 日

# 厦门大学学位论文著作权作用声明

本人完全了解厦门大学有关保留、使用学位论文的规定。厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅，有权将学校论文的内容编入有关数据库进行检索，有权将学校论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

本学位论文属于

- 1、保密（ ），在 年解密后适用本授权书。
- 2、不保密（ ）

（请在以上相应括号内打“√”）

作者签名：

日期： 年 月 日

导师签名：

日期： 年 月 日

厦门大学博硕士学位论文摘要库

## 内 容 摘 要

数据挖掘 (Data Mining) 就是近年来随着人工智能和数据库技术的发展而出现的一门新兴学科。关联规则 (Association Rule) 是数据挖掘的重要的研究课题, 是数据挖掘的主要技术之一, 目的就是要在交易数据库中发现各项目之间的关系。它在决策支持系统、专家系统和智能型信息系统等各个方面都起着重要的作用。近年来有许多学者都在进行关联规则的研究, 并在原有的方法基础上进行了很大的改进, 还结合不同领域的知识得出了多种不同的关联分析方法。

本文对基本关联规则的现状进行了剖析, 指出了其两个重要衡量工具—支持度和置信度的表达方式上的混淆之处, 并从集合论和概率论两个角度给出了规范化描述, 为以后的研究分析奠定了基础。

随后采用灰色关联分析进行实证分析。大多数学者都是将灰色关联规则用于研究方案的可行性, 即与参考方案的相似程度, 而本文将该方法用于分析调查问卷, 就问卷调查结果分析了问卷中各问题的关联度。

然后在现有的模糊集关联规则研究的基础上, 结合数据统计特征提出了具体的隶属函数, 再在灰色关联规则分析的基础上, 进行模糊关联规则分析。

最后, 将灰色关联规则和模糊关联规则结合在一起, 将灰色关联系数作为权数加入到模糊关联规则中, 进行加权模糊关联规则分析, 并以厦门市的产业结构作为实证分析的对象进行了研究, 并就分析结果对厦门市的产业结构调整提出了几点建议。

本文的创新点主要体现在以下几个方面: 第一, 在总结现有关联规则的表达方式的基础上, 指出了其中的矛盾以及容易产生的混淆, 并进行了规范化描述; 第二, 结合原始数据的特征, 提出具体的隶属函数形式; 第三, 将两种关联规则结合起来, 以灰关联度作为权数加入到模糊关联分析中。

**关键词:** 数据挖掘; 关联规则; 灰色理论; 模糊集

厦门大学博硕士学位论文摘要库

## Abstract

Data mining is a new subject which arises with the development of artificial intelligence and database. Association rule is one of the important research tasks, and also one of the key technologies in data mining. Its purpose is to find out the association between all items in business database. Association rule plays a very important role in decision system, expert system and aptitude information system in every aspect. Many experts have been making research on association rule, and have made big progress based on original methods. They also find out several new methods with respect to the knowledge of many other fields.

This paper gives out an analysis of the definitions of association rule in the existing books and papers. Then it points out the cause of the difference, normalizes the description and makes out new descriptions from the angle of set and possibility, which lays a foundation on the further research of the theory and application of the association rule.

Then we use the grey association rule to make empirical analysis. This theory is often used to analyze the feasibility of a project, but in this paper it is used to find out the relationship between the questions in a questionnaire by analyzing the results.

After that, a membership function is given out by considering the feature of the original data, which is used to analyze the result given by the grey association rule.

In the end, we use the two rules together to analyze the industrial structure of Xiamen City. The grey association coefficient is used as part of the weight in fuzzy association rule to find out fuzzy weighted association rule. Finally we give out some suggestions according to the analysis result.

The main contributions of this dissertation are that firstly, we point out the inconsistency and the possible confusion that it leads to based on the summarization of the existing expression of association rule, and normalize the description ; secondly,

厦门大学博硕士学位论文摘要库



gives out a membership function considering the feature of data; thirdly, we link the two methods together, the grey association coefficient is used to be part of the weight in fuzzy weighted association rule.

**Key words :** data mining; association rule; grey theory; fuzzy set

厦门大学博硕士学位论文摘要库

厦门大学博硕士学位论文摘要库

## 目 录

<b>第一章 引言</b> .....	<b>1</b>
第一节 数据挖掘综述 .....	1
第二节 关联规则概述 .....	4
<b>第二章 关联规则的规范化描述</b> .....	<b>7</b>
第一节 关联规则概念的剖析 .....	7
第二节 关联规则概念的规范 .....	10
<b>第三章 灰色关联分析</b> .....	<b>14</b>
第一节 灰色系统理论的引入 .....	14
第二节 灰色理论 .....	14
第三节 灰色关联分析及其模拟 .....	17
<b>第四章 模糊关联规则</b> .....	<b>21</b>
第一节 关联规则中模糊概念的引入 .....	21
第二节 模糊集方法 .....	22
第三节 模糊关联规则的算法 .....	25
第四节 加权模糊关联规则 .....	31
<b>第五章 厦门市产业结构的实证分析</b> .....	<b>33</b>
第一节 我国产业结构的三次重大调整 .....	33
第二节 厦门市产业结构关联分析 .....	35
第三节 厦门市产业结构调整的对策 .....	46
<b>参考文献</b> .....	<b>48</b>
<b>后 记</b> .....	<b>50</b>

厦门大学博硕士学位论文摘要库

## Abstracts

<b>Chapter 1 Introduction .....</b>	<b>1</b>
1.1 Summerization of Data Mining .....	1
1.2 Summerization of Association Rule .....	4
<b>Chapter 2 The normalize description of Association rule .....</b>	<b>7</b>
2.1 Analysis of the description of Association Rule .....	7
2.2 Normalization of Association Rule .....	10
<b>Chapter 3 Gray Association Rule .....</b>	<b>14</b>
3.1 The Import of Gray System Theory .....	14
3.2 The Gray Theory .....	14
3.3 Gray Association Analysis and its simulation .....	17
<b>Chapter 4 Fuzzy Association Rule .....</b>	<b>21</b>
4.1 The Import of Fuzzy set in Association Rule .....	21
4.2 Techniques of Fuzzy Set .....	22
4.3 Algorithm of Fuzzy Association Rule .....	24
4.4 Weighted Fuzzy Association Rule .....	31
<b>Chapter 5 Empirical Analysis of the industrial Structure of Xiamen City .....</b>	<b>33</b>
5.1 Three Important adjustments of the Industrial Structure in China.....	33
5.2 Association Analysis of the Industrial Structure of Xiamen City .....	35
5.3 Suggestions about the Adjustment of the Industrial Structure of Xiamen City .....	46
<b>Reference .....</b>	<b>48</b>
<b>Postscript .....</b>	<b>50</b>

厦门大学博硕士学位论文摘要库

## 第一章 引言

### 第一节 数据挖掘综述

#### 一、什么是数据挖掘

随着科学技术的不断进步和全球化的向前推进，世界各地的交流已经趋于频繁和便利，Internet技术的出现和发展已将整个世界连接在一起，人们可以穿越时空般地在网上交换信息和工作，在这个信息爆炸的时代，面对着伴随出现的大量数据信息，如何去粗取精，去伪存真将浩如烟海的数据转换成知识的技术是一个急需解决的问题。数据挖掘（Data Mining）就是近年来随着人工智能和数据库技术的发展而出现的一门新兴学科。它是通过仔细分析大量的数据，从中筛选出隐含的、可信的、新颖的、有效的信息的高级处理过程。它是一门交叉性学科，融合了人工智能、数据库技术、模式识别、机器学习、统计学和数据可视化等多个领域的知识和技术。

#### 二、数据挖掘的主要对象

数据挖掘的范围非常广泛，可以是社会科学、经济学、商业数据、科学处理产生的数据和卫星观测得到的数据。它们的数据结构也各不相同，可以是层次的、网状的、关系的和面向对象的数据。

关系数据库是表的集合，每个表都赋予一个唯一的名字。每个表包含一组属性（列或字段），并通常存放大量元组（记录或行）。关系中的每个元组代表一个被唯一的關鍵字标识的对象，并被一组属性值描述。

当数据挖掘用于关系数据库时，可以进一步搜索趋势或数据模式。数据挖掘系统也可以检测偏差，如在商业营运中，与以前的年份相比，哪种商品的销售出人预料。这种偏差可以进一步考察，例如包装是否有变化，或价格是否大幅度提高。

随着数据库技术的发展，各种高级数据库系统已经出现并在开发中，以适应新的数据库应用需要。

新的数据库应用包括处理空间数据(如地图)、工程设计数据、超文本和多媒体数据、时间相关的数据(如历史数据或股票交易数据)和Web。

为响应这些需求,开发了高级数据库系统和面向特殊应用的数据库系统。这些包括面向对象和对象-关系数据库系统、空间数据库系统、时间和时间序列数据库系统、文本和多媒体数据库系统、异种和遗产数据库系统、基于Web的全球信息系统。

事务数据库由一个文件组成,其中每个记录代表一个事务。通常,一个事务包含一个唯一的事务标识号,和一个组成事务的项的列表(如,在商店购买的商品)。事务数据库可能有一些与之相关联的附加表,包含关于销售的其他信息,如事务的日期、顾客的ID号、销售者的ID号、销售分店等等。

### 三、数据挖掘的主要任务

数据挖掘的产生来源于需求。随着人类社会步入信息时代,信息的获得和利用在生产 and 生活中占据越来越重要的地位。作为信息载体或者知识载体的数据,其数量空前地巨大。日益成熟的数据库系统和数据库管理系统,以计算机网络技术为代表的数据的传输和远程交互技术手段,使得人们生成、采集和传输数据的能力都有了巨大增长,而数据挖掘的任务,则是在此基础上完成从原始数据到信息和知识的转化。数据挖掘的任务可以分为以下几个方面:

1. 总结规则的挖掘,就是从客户指定的数据中挖掘出(从不同角度或在不同的层次上)平均值/极小值/极大值、总和、百分比等。
2. 关联规则挖掘,即从客户指令的数据库中挖掘出满足一定条件的依赖性关系。
3. 分类规则挖掘,也就是在已知训练信息的特征和分类结果的基础上,为每一类别找到一个合理的描述或模型,然后再用这些分类的描述或模型来对未知的新数据进行分类。
4. 群集规则挖掘是根据事物本身的特性,研究个体分类的方法,目的在于客观地按被处理对象的特征分类,有相同特征的对象归为一类。与分类规则挖掘的区别在于群集不需要事先定义好该如何分类,同时也不需要训练组的数据。
5. 预测分析,这是在当分类工作偏向于插入漏掉的数据、预测数据分类或发



Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to [etd@xmu.edu.cn](mailto:etd@xmu.edu.cn) for delivery details.

厦门大学博硕士论文摘要库