

摘要

AFS 理论是刘晓东教授于 1995 年首先提出的;它是 Axiomatic Fuzzy Sets 的缩写,即公理模糊集;能被用来研究人类思维的法则,便于计算机操作,较好地揭示人靠经验和直觉描述复杂的模糊概念以及确定相应模糊概念的隶属函数的内在机理。AFS 理论是从问题的原始数据出发,用 AFS 结构、AFS 代数的运算和其上的一个逆序对合运算来建立模糊逻辑系统,用拓扑分子格刻划人类概念之间的抽象关系,使得隶属函数和模糊逻辑系统的建立更具客观性、严密性和统一性。AFS 理论已经尝试着被应用于聚类分析、模式识别和故障诊断等领域。

本文对 AFS 模糊逻辑的聚类分析算法(X. D. Liu, W. Wang and T. Y. Chai. IEEE Transaction on Systems, Man, Cybernetics, 2005)进行了进一步的研究,对其置信度算法进行了改进,并将改进后的指标应用到含有 150 个样本的 Iris 数据(见 <ftp://ftp.ics.uci.edu/pub/machine-learning-databases/Iris/>)。本文的算法只用到了样本属性上的序关系。本研究表明只用样本属性上的序关系,AFS 模糊逻辑分类分析算法也能够获得很高的准确率,因此此算法能够很好的应用到那些样本属性只能用序关系描述而无法用数值描述的数据集。并且对 X.D.Liu,W.Wang and T.Y.Chai 定义的模糊聚类指标进行修改,并做进一步的研究,同时在 X.D.Liu, W.Wang and T.Y.Chai 提出的 AFS 模糊逻辑聚类分析算法的基础上设计了一种新的分类器,并将它应用于 Iris 数据。最后利用 Iris 数据对分类器的稳定性和准确性进行了研究,这些研究表明这种新分类器简单、有效。

关键词: AFS 理论; EI 代数; *EI 代数; 模糊聚类分析; 模糊分类器

Fuzzy Classification Approaches Based on

AFS Fuzzy Logic

Abstract

AFS theory was proposed first by professor Liu Xiaodong in 1995. The meaning of AFS is Axiomatic Fuzzy Sets. It can be used to study the law of human thinking and be easily operated by computers. It makes some of the mechanisms of decompose and composition of human conceptions to be understood with mathematical terms. Also it preferably clarified both the essential mechanism that humans use to describe complex fuzzy concepts that depend upon experiences and intuition and the essential mechanism that humans use to choose the corresponding membership function. AFS theory embodies the abstract relationship between human conceptions using topological molecular lattices. It makes the set up of membership functions and fuzzy logic system full of more objectivity, rigor and unity, All of these result come from the original data .So much theories of AFS has already tried to be applied on clustering, pattern recognition and diagnosis ,etc.

In this paper, the AFS Fuzzy logic clustering algorithm (X.D. Liu, W.Wang and T.Y.Chai.IEEE Transaction on Systems, Man, Cybernetics, 2005) have been studied further by the improvement of the algorithm and the application of the algorithm to Iris data (reference <ftp://ftp.ics.uci.edu/pub/machine-learning-databases/Iris/>). In stead of example of less than 10 samples, we apply the improved algorithm to Iris data which has 150 samples and just the order relationship of the samples on the attributes is used. This study shows that the AFS Fuzzy logic clustering algorithm can obtain much higher reclassification accuracy according to the order relationship. Thus the algorithm can be applied to the data sets in which the attributes are only described by order relationship. And improve and study the fuzzy clustering index defined by X.D.Liu, W.Wang and

T.Y.Chai, and design a new classifier based on the AFS fuzzy logic clustering algorithm proposed by X.D.Liu, W.Wang and T.Y.Chai. As an example, the new classifier is applied to Iris data (<ftp://ftp.ics.uci.edu/pub/machine-learning-databases/Iris/>). The stability and reclassification accuracy of the classifiers are studied through the Iris data, the examples show that the new classifiers are simple and efficient.

Key Words: *AFS* theory; *EI* algebras; **EI* algebras; Fuzzy Cluster Analysis; Fuzzy Classifier.

大连海事大学学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明：本论文是在导师的指导下，独立进行研究工作所取得的成果，撰写成硕士学位论文“基于 AFS 模糊逻辑的分类分析方法研究”。除论文中已经注明引用的内容外，对论文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本论文中不包含任何未加明确注明的其他个人或集体已经公开发表或未公开发表的成果。

本声明的法律责任由本人承担。

论文作者签名： 理昌 07年3月25日

学位论文版权使用授权书

本学位论文作者及指导教师完全了解“大连海事大学研究生学位论文提交、版权使用管理办法”，同意大连海事大学保留并向国家有关部门或机构送交学位论文的复印件和电子版，允许论文被查阅和借阅。本人授权大连海事大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，也可采用影印、缩印或扫描等复制手段保存和汇编学位论文。

保密 ，在 _____ 年解密后适用本授权书。

本学位论文属于： 保密
不保密 (请在以上方框内打“√”)

论文作者签名： 理昌 导师签名： 刘峰东

日期： 07 年 3 月 25 日

引 言

模糊性是人类认识和思维的重要特征之一。自 1965 年, L. A. Zadeh 提出模糊集理论以来, 很多数学家、工程师和技术人员应用它取得了许多重要的结论。国内外许多学者在理论方面做了大量的研究, AFS 理论就是其中之一。

自 1995 年 AFS 理论被提出来, 刘晓东教授在 AFS 理论的理论和应用两方面都做了大量的工作 (见文献[10-18], [42-48])。AFS 理论给出了三个新的数学对象: AFS 代数、AFS 结构和认知域, 对模糊数学的理论和应用都有很大的促进作用, 特别是在应用方面, AFS 理论已被尝试着应用于聚类分析、模式识别和故障诊断等领域。研究成果多数发表在国内外重要的学术期刊上, 已初步得到国内外专家学者的认可和关注。

本文在 AFS 的理论理解、发展和应用方面都做了一定的研究, 下面详细展开。

第 1 章 模糊数学基础

模糊数学理论虽然经历了近三十年的发展，但是相对于其他数学学科，它还是一门新的数学理论、一种新的数学工具。早在 70 年代初，模糊集合的概念逐渐为更多的学者所关注。有关模糊集的论文数量迅速增加，质量不断提高，涉及的范围十分广泛。模糊理论与经典理论相比，更接近于人类思维和自然语言系统，因此更适合于模仿人类的思维推理。

在日常生活中，经常遇到许多模糊事物，没有分明的数量界限，要使用一些模糊的词句来形容、描述。比如：年轻、高个、漂亮……，在人们的工作经验中，往往也有许多模糊的东西。例如，要确定一炉钢水是否已经炼好，除了要知道钢水的温度、成分比例和冶炼时间等精确信息外，还需要参考钢水颜色、沸腾情况等模糊信息。因此，除了很早就有涉及误差的计算数学之外，还需要模糊数学。

与计算机相比，人脑具有处理模糊信息的能力，善于判断和处理模糊现象。但计算机对模糊现象识别能力较差，为了提高计算机识别模糊现象的能力，就需要把人们常用的模糊语言设计成机器能接受的指令和程序，以便机器能像人脑那样简洁灵活的做出相应的判断，从而提高自动识别和控制模糊现象的效率。这样，就需要寻找一种描述和加工模糊信息的数学工具，这就推动数学家深入研究模糊数学。所以，模糊数学的产生是有其科学技术与数学发展的必然性。

本章主要是简单介绍模糊数学的基础知识（文献[1-9]）。

1.1 模糊集理论的建立

模糊性的产生源于模糊概念，所谓模糊概念是指有一定内涵但没有明确外延的概念。例如，“青年”这一概念，即哪个年龄段的人是青年，很难说清楚，这就是一个模糊概念。换句话说，模糊性就是指客观事物差异的不分明性。随着社会的发展，过去那些与数学毫无关系或关系不大的学科，都迫切需要定量化和数学化，从而遇到大量的模糊概念，这也正是由这些学科本身的特点所决定的。人们不能为迁就现有的数学方法而改变出这些学科决定的客观规律，而只能改造数学，使它的应用更为广泛，模糊数学就是在这样的背景下形成的。1965 年，美国加州大学的 L. A. Zadeh 教授发表了题为“Fuzzy Sets”的论文，奠定了模糊集理论及

应用研究的基础,标志着模糊数学的诞生。他在论文“Fuzzy Sets”中,建立了模糊集合论,引入了“隶属函数”来描述差异的中间过渡,这是精确性对模糊性的一种逼近。隶属函数的引入标志着模糊数学的建立,因为它是描述模糊性的关键。L. A. Zadeh 教授首次成功地运用了数学方法描述模糊概念,这无疑是一项开创性的工作。

1.2 模糊数学基本概念

1.2.1 模糊集、隶属度、隶属度函数及模糊逻辑

模糊集是一种边界不分明集合,模糊集与普通集合既有区别又有联系。对于普通集合而言,任何一个元素要么属于该集合,要么不属于该集合,非此即彼,具有精确明了的边界;而对于模糊集合,一个元素可能具有既属于该集合的成份又具有不属于该集合的成份,亦此亦彼,边界不分明或界限模糊。

模糊集使得某元素可以以一定程度属于某集合,某元素属于某集合的程度由“0”与“1”之间的一个数值——隶属度来刻画或描述。把一个具体的元素映射到一个合适的隶属度是由隶属函数来实现的。隶属函数可以是任意形状的曲线,取什么形状取决于是否让我们使用起来感到简单、方便、快速、有效。唯一的约束条件是隶属度函数的值域为 $[0, 1]$ 。现有的隶属函数都是主观人为给出的。

模糊逻辑是建立在模糊集基础上的,任何陈述或命题的真实性只是一定程度的真实性,与建立在普通集合基础上的布尔逻辑相比,模糊逻辑是一种广义化的逻辑。在布尔逻辑中,任何陈述或命题只有两种取值,即逻辑真和逻辑假,常用“1”表示逻辑真,“0”表示逻辑假,而在模糊逻辑中,陈述或命题的取值除真和假(“1”和“0”)外,可取“0”与“1”之间的任何值,如 0.75,即命题或陈述在多大程度上为真或假,例如“老人”这一概念,在普通集合中需要定义一个明确的边界,如 60 岁以上是老人,而在模糊集合,老人的定义集合没有一个明确的边界,60 岁以上是老人,58 岁也属于老人,40 岁在一定程度上也属于老人,只是他们属于老人这一集合的程度不同而已。

1.2.2 模糊集的格运算

在偏序集 (L, \leq) 中,对任意 $x, y \in L$,若 $x \vee y, x \wedge y$ 均存在,则称 (L, \leq) 为格。

在格 (L, \leq) 中, 三个条件 $x \leq y$, $x \vee y = y$, $x \wedge y = x$ 是相互等价的。若将“ \vee ”、“ \wedge ”看成是在集 L 上定义的二元运算, 则这二种运算具有如下性质:

- (1) 交换律 $x \vee y = y \vee x$, $x \wedge y = y \wedge x$;
 (2) 结合律 $(x \vee y) \vee z = x \vee (y \vee z)$, $(x \wedge y) \wedge z = x \wedge (y \wedge z)$;
 (3) 吸收律 $x \vee (x \wedge y) = x$, $x \wedge (x \vee y) = x$;

因此, 格 (L, \leq) 亦称为具有“ \vee ”、“ \wedge ”两种运算的代数系统, 有时记为 (L, \leq, \vee, \wedge) , 或简记为 (L, \vee, \wedge) 。

若格 (L, \leq) 中, 对任 $A \subseteq L$, 都存在 $\sup A$ 及 $\inf A$, 则称 (L, \leq) 为完全格(或完备格)。

若格 (L, \leq, \vee, \wedge) 满足分配律 $x \vee (y \wedge z) = (x \vee y) \wedge (x \vee z)$, $x \wedge (y \vee z) = (x \wedge y) \vee (x \wedge z)$ 则称之为分配格。

在格 (L, \leq, \vee, \wedge) 中, 若存在最大元 1 及最小元 0 , 且对任 $x \in L$ 存在 $y \in L$, 使 $x \vee y = 1$, $x \wedge y = 0$, 则称 (L, \leq) 为有补格, 并称 y 为 x 的补元, 记为 \bar{x} 。

若格 (L, \leq) 是有补格和分配格, 则称 (L, \leq) 为布尔格(布尔代数)。在布尔格中, 对每个元素来说, 只有一个补元。

1.2.3 模糊关系

对于集合 X, Y , 直积 $X \times Y$ 的子集 R 称为 X 与 Y 之间的二元关系, 简称关系。若 $X=Y$, 则 $X \times Y$ 的子集 R 称为 X 上的二元关系。若 $(x, y) \in R$, 则称“ x 与 y 具有关系 R ”, 也记作 xRy 。若 $(x, y) \notin R$, 则记作 $x\bar{R}y$ 。

集合 X 上的几个重要的二元关系:

- (1) 自反关系 R $\forall x \in X$ 恒有 xRx
 (2) 对称关系 R 若 xRy 则 yRx
 (3) 反对称关系 R 若 xRy 且 yRx , 则 $x=y$
 (4) 传递关系 R 若 xRy 且 yRz , 则 xRz

具有自反、对称、传递三种性质的关系叫等价关系。由等价关系 R 可定义集合 $[x] = \{y | xRy\}$, 称 $[x]$ 为 x 的等价类。

具有自反、反对称、传递三种性质的关系 R 称为偏序关系。通常将偏序关系 R 记成 “ \leq ”。

给了某个偏序关系 “ \leq ” 的集 X , 称为偏序集, 记为 (X, \leq) 。

若 (X, \leq) 是半序集, 且对任 $x, y \in X$, 必有 $x \leq y$ 或 $y \leq x$, 则称 (X, \leq) 为全序集, 这时称 “ \leq ” 为 X 上的全序。

设 (\varnothing, \leq) 是半序集, 若存在 $a \in \varnothing$, 使对任意 $x \in \varnothing$ 有 $x \leq a$ (或 $\geq a$), 则称 a 为 \varnothing 的最大(小)元。

设 (\varnothing, \leq) 是半序集, $a \in \varnothing$ 。若存在 $a \in \varnothing$, 使对 A 中任意元 x , 均有 $x \leq a$ (或 $\geq a$), 则称 a 为 A 的上界(下界)。上界中的最小元素称为 A 的上确界, 记为 $\sup A$; 下界中的最大元素称为 A 的下确界, 记为 $\inf A$ 。对于 \varnothing 中的元素 x, y 来说, 若 x, y 的上确界(下确界)存在, 则记为 $x \vee y$ ($x \wedge y$)。

1.3 模糊集合理论

1.3.1 模糊集合的运算

定义模糊集合的运算方法, 与定义普通集合运算方法一样。运算所得的新模糊集合的隶属函数是利用参加运算的模糊集合的隶属函数来定义的。两模糊集合的具体运算, 实际上就是逐点的对隶属度作相应的运算。以下归纳了论域 X 中的某些模糊集合运算。

- 集合相等 $A=B \Leftrightarrow \mu_A(x) = \mu_B(x)$ 对所有 $x \in X$
- 集合补 $A' \Leftrightarrow \mu_{A'}(x) = 1 - \mu_A(x)$ 对所有 $x \in X$
- 集合包含 $A \subseteq B$ 模糊集合 A 包含于 B 或是 B 的子集 \Leftrightarrow
 $\mu_A(x) \leq \mu_B(x)$ 对所有 $x \in X$
- 集合并 $A \cup B \Leftrightarrow \mu_{A \cup B}(x) = \mu_A(x) \vee \mu_B(x)$ 对所有 $x \in X$
 其中连接运算符 \vee 表示参量中最大的。
- 集合交 $A \cap B \Leftrightarrow \mu_{A \cap B}(x) = \mu_A(x) \wedge \mu_B(x)$ 对所有 $x \in X$
 其中相交运算符 \wedge 表示参量中最小的。
- 集合积 $AB \Leftrightarrow \mu_{AB}(x) = \mu_A(x) \mu_B(x)$ 对所有 $x \in X$

- 集合的幂 $A^N \Leftrightarrow \mu_{A^N}(x) = (\mu_A(x))^N$
- 概率和 $A \hat{+} B \Leftrightarrow \mu_{A \hat{+} B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x)\mu_B(x) = 1 - (1 - \mu_A(x))(1 - \mu_B(x))$
其中，“ \vee ”和“ \wedge ”是普通算术运算符。
- 有界和 $A \oplus B \Leftrightarrow \mu_{A \oplus B}(x) = \min(1, (\mu_A(x) + \mu_B(x)))$
- 有界积 $A \beta B \Leftrightarrow \mu_{A \beta B}(x) = \max(0, (\mu_A(x) + \mu_B(x) - 1))$
- 有界差 $A | - | B \Leftrightarrow \mu_{A | - | B}(x) = \max(0, (\mu_A(x) - \mu_B(x)))$
其中，分隔 μ_A 和 μ_B 的是算术差运算符， $A | - | B$ 代表那些 A 中有而 B 中无的元素。
- 标准化 $\text{NORM}(A) \Leftrightarrow \mu_{\text{NORM}(A)}(x) = \mu_A(x) / \max\{\mu_A(x)\}$
其中， $\max()$ 返回所有元素的隶属度的最大值。如果最大值小于1，那么所有的隶属度都会增加，如果 $\max=1$ ，那么隶属度不变。

1.3.2 模糊数学的发展及应用

模糊数学是一门新兴学科，自其诞生以来，发展速度非常快，研究领域越来越宽广。它已初步应用于模糊控制、模糊识别、模糊聚类分析、模糊决策、模糊评判、系统理论、信息检索、控制、心理学等各个方面。它的研究领域可大体分为三个方面：

- (1)模糊数学理论，含与经典数学、统计数学的关系；
- (2)模糊语言、模糊逻辑、模糊控制的研究；
- (3)模糊数学在自然科学及社会科学中的应用。

到目前为止，在自然科学、社会科学、工程技术的各个领域，都会涉及大量的模糊因素和模糊信息处理问题，模糊技术几乎渗透到了所有领域。像冶金、机械、石油、化工、电力、电子、轻工、能源、交通、医疗、卫生、农业、林业、地理、水文、地震、气象、环保、建筑、行为科学、管理科学、法学、教育、军事科学等等，每个领域都有其成功的应用范例。

第 2 章 数据挖掘概况

2.1 什么是数据挖掘

随着社会的进步,数据库技术的不断发展及数据库管理系统的广泛应用,数据库中存储的数据量急剧增大。在大量的数据背后隐藏着许多重要信息,人们根据各自的需要都希望从中得到有用的信息,而这些重要信息可以很好地支持人们的决策。但人们通过这些数据所获得的信息量仅仅是这个数据库所包含信息量的一部分,隐藏在这些数据之后的更重要信息是关于这些数据的整体特征的描述以及对其发展趋势的预测,这些信息在决策生成的过程中具有重要的参考价值。在“寻找”这些信息方面中外学者作了大量的研究工作(见文献[19-28]),下面简单介绍他们的成果。

2.1.1 数据挖掘的定义

数据挖掘是从大量数据中挖掘出隐含的、先前未知的、对决策有潜在价值的信息或知识和规则的过程。这些规则蕴含了数据库中一组对象之间的特定关系,揭示出一些有用的信息,为经营决策、市场策划、金融预测等提供依据。简单地说,数据挖掘是从大量数据中提取或“挖掘”知识。

数据挖掘的目的是帮助分析人员寻找数据之间的关联,发现被忽略的要素,而这些信息对于预测趋势和决策行为是十分有用的。例如:从超市的大量事务数据中发现,许多人在购买计算机的同时要购买价格合适的打印机,于是将这两种商品放在很近的位置,方便顾客购买,结果这两种商品的销售量都增加了许多。

通过数据挖掘,有价值的知识、规则或高层次的信息就能从数据库的相关数据集集中抽取出来,并从不同角度显示,从而使大型数据库作为一个丰富可靠的资源为知识归纳服务。

2.1.2 数据挖掘的分类

由于数据挖掘源于多个学科,因此数据挖掘研究就产生了大量的、各种不同类型的数据挖掘系统。这样,就需要对数据挖掘系统给出一个清楚分类。这种分类可以帮助用户区分数据挖掘系统,确定最适合其需要的数据挖掘系统。根据

不同的标准，数据挖掘系统可以分类如下：

根据挖掘的任务分类有：分类或预测模型知识发现、数据总结、数据聚类、关联规则发现、序列模型发现、依赖关系或依赖模型发现、异常和趋势发现等等。

根据挖掘的数据库类型分类有：关系数据库、数据仓库、事物数据库、高级数据库系统和高级数据库应用（包括面向对象数据库、空间数据库、时态数据库、文本数据源、多媒体数据库、异质数据库、遗产数据库，以及环球网等）。

根据挖掘方法分类有：统计方法、机器学习方法、神经网络方法和数据库方法。统计方法中，可细分为：回归分析（多元回归、自回归等）、判别分析（贝叶斯判别、费歇尔判别、非参数判别等）、聚类分析（系统聚类、动态聚类等）、探索性分析（主元分析、相关分析法等）等。

2.2 数据挖掘的功能

数据挖掘功能用于指定数据挖掘任务中要找的模式类型。数据挖掘任务一般可以分为两大类：描述和预测。描述性挖掘任务是刻画数据库中数据的一般特性；预测性挖掘任务在当前数据上进行推断，以进行预测。预测是构造和使用模型评估无标号样本类，或评估给定样本可能具有的属性值或区间值。预测的目的是从历史数据纪录中自动推导出给定数据的推广描述，从而能对未来数据进行预测。

数据挖掘功能以及他们的模式类型介绍如下：

2.2.1 分类

分类在数据挖掘中是一项非常重要的任务。其方法是找出描述并区分数据类或概念的模型（或函数），以便能够使用模型预测类标记未知的对象类；其目的是学会一个分类函数或分类模型（也常常称作分类器），该模型能把数据库中的数据项映射到给定类别中的某一个。

分类器的构造方法有统计方法、机器学习方法、神经网络方法和粗糙集等等。主要方法有决策树、规则归纳、面向大量数据的分类知识发现技术。

2.2.2 数据聚类分析

将物理或抽象对象的集合分组成为由类似的对象组成的多个类的过程被称为聚类。聚类是把一组个体按照相似性归成若干类别，即“物以类聚”。它的目的是

使得属于同一类别的个体之间的距离尽可能的小而不同类别上的个体间的距离尽可能的大。聚类方法包括统计方法、机器学习方法、神经网络方法和面向数据库的方法。

在统计方法中, 聚类称聚类分析, 它是多元数据分析的三大方法之一(其他两种是回归分析和判别分析)。它主要研究基于几何距离的聚类, 如欧式距离、明考斯基距离等。传统的统计聚类分析方法包括系统聚类法、分解法、加入法、动态聚类法、有序样品聚类、有重叠聚类和模糊聚类等。

在数据挖掘中, 面临的常常是含有大量数据的数据库, 因而要探讨面向数据库的聚类方法以适应新问题带来的挑战。关于大数据库的聚类问题近来已在数据库研究界得到了相当的重视。聚类技术主要有: 划分方法、层次方法、基于密度的方法, 基于网络的方法和基于模型的方法。

2.2.3 关联分析

关联分析即发现关联规则, 关联规则是形式如下的一种蕴含或规则: $A \Rightarrow B$, 即 " $A_i \cap \dots \cap A_m \Rightarrow B_j \cap \dots \cap B_n$ ", 其中, $A_i (i \in \{1, \dots, m\})$, $B_j (j \in \{1, \dots, n\})$ 是属性一值对。关联规则 " $A \Rightarrow B$ " 解释为 "满足 A 中条件的数据库元组多半也满足 B 中条件"。例如: "72% 包含 Item A, B 和 C 的记录同时也包含 Item D 和 E"。其中 72% 称为置信度。

规则 $A \Rightarrow B$ 在事物集 D 中成立, 具有支持度 s , 其中 s 是 D 中事物包含 $A \cup B$ (即 A 和 B 两者) 的百分比。它是概率 $p(A \cup B)$ 。规则 $A \Rightarrow B$ 在事物集 D 中具有置信度 c , 如果 D 中包含 A 的事物同时也包含 B 的百分比是 c 。这是条件概率 $p(B|A)$ 。

同时满足最小支持度阈值和最小置信度阈值规则称作强规则。

关联规则有多种分类方法, 如: 根据规则中所处理值的类型、涉及的数据维、涉及的抽象层和根据关联挖掘的各种扩充等。用于关联规则发现的对象主要是事物型数据库。一个事物一般有如下几个部分组成: 事物处理时间, 一组顾客购买的物品, 有时也有顾客标识号(如信用卡号)。

计算关联规则的关键是如何高效地求出高频物品集。目前在关联规则发现的研究论文中, 提出的好几种方法都是围绕此问题而来的。典型的算法有 Apriori 和 DHP 算法。

2.2.4 概念描述

数据可以与类或概念相关联。用汇总的、简洁的、精确的方式描述每个类和概念可能是有用的。这种类和概念的描述称为类/概念描述。这种描述可以通过下述方法得到：1)数据特征化，2)数据区分，3)数据特征化和比较。

数据特征化是目标类数据的一般特征或特性的汇总。通常，用户指定类的数据通过数据库查询收集。数据区分是将目标类对象的一般特性与一个或多个对比类对象的一般特性比较。目标类和对比类由用户来指定，而对应的数据通过数据库查询检索。

2.2.5 孤立点分析

数据库中可能包含一些数据对象，它们与数据的一般行为或模型不一致。这些数据对象是孤立点。大部分数据挖掘方法将孤立点视为噪声或异常而丢失。然而，在一些应用中(如欺骗检测)，罕见的事件可能比正常出现的那些更有趣。孤立点数据分析称作孤立点挖掘，孤立点可以使用统计试验检测。

2.2.6 演变分析

数据演变分析是用来描述随时间变化的对象的规律或趋势，并对其建模。尽管它可能包括时间相关数据的特征化、区分、关联、分类或聚类，其特点包括时间序列数据分析、序列或周期模式匹配和基于类似性的数据分析。

2.3 数据挖掘的应用

数据挖掘技术被越来越多的领域所采用，并取得了较好的效果，为人们的正确决策提供了很大的帮助。下面简单介绍它在科研和实践中的应用。

2.3.1 在科学研究中的应用

在过去的十年里，生物医学研究有了迅猛的发展，针对生物学和 DNA 数据分析的数据挖掘开始被使用。目前生物医学的大量研究都集中在 DNA 数据的分析上。数据挖掘成了 DNA 数据分析的强有力工具。

数据挖掘在天文学上也有着广泛的应用。有一个非常著名的应用系统：SKICAT(Sky Image Cataloging and Analysis Tool)。它是美国加州理工学院喷气推进实验室与天文科学家合作开发的用于帮助天文学家发现遥远的类星体的一个工

具。利用它，天文学家已发现了 16 个新的极其遥远的类星体，该项发现能帮助天文工作者更好地研究类星体的形成及其早期宇宙的结构。

其实，数据挖掘在科研中的应用非常广泛，这里只是提到几个典型的例子。

2.3.2 在实践中的应用

数据挖掘在实践中的应用也非常广泛，下面仅举几个典型的例子说明。

在金融投资方面，典型的金融分析领域有投资评估和股票交易市场预测，分析方法一般采用模型预测法。由于金融投资的风险很大，在进行投资决策时，更需要通过对各种投资方向的有关数据进行分类，以选择最佳的投资方向。大部分银行和金融机构都提供丰富多样的储蓄服务，信用服务和投资服务。在银行和金融机构中产生的金融数据通常相对比较完整、可靠和高质量，这就很适合于系统化的数据分析和数据挖掘。

在市场营销方面，由于管理信息系统和 POS 系统在商业尤其是零售业内的普遍使用，特别是条形码技术的使用，从而可以收集到大量关于用户购买情况的数据，并且数据是在不断激增。利用数据挖掘技术通过对用户数据的分析，可以得到关于顾客购买取向和兴趣的信息，从而为商业决策提供了可靠的依据。

在产品制造方面，随着现代技术越来越多的应用于产品制造，制造业已不是人们想象中的手工劳动，而是集成了多种先进科技的流水作业。在产品的生产制造过程中常常伴随有大量的数据，如产品的各种加工条件或控制参数(如时间，温度等)，这些数据反映了每个生产环节的状态，不仅为生产的顺利进行提供了保证，而且通过对这些数据的分析，得到产品质量与这些参数之间的关系。这样通过数据挖掘对这些数据的分析，可以对改进产品质量提出针对性很强的建议，而且有可能提供出新的更高效节约的控制模式，从而为制造厂家带来极大的回报。

在电信业已经迅速地单纯的提供市话和长话服务演变为提供综合电信服务，如语音，传真，寻呼，移动电话，图像，电子邮件，计算机和 Web 数据传输，以及其他数据通信服务。利用数据挖掘可以改进电信服务，主要体现在如下几个方面：(1)电信数据的多维分析；(2)盗用模式分析和异常模式分析；(3)多维关联和序列模式分析；(4)电信数据分析中可视化工具的使用。

2.4 数据挖掘的一般步骤

首先是数据预处理。存在不完整的、含噪声的和不一致的数据是大型的现实世界数据库或数据仓库的共同特点。对数据进行预处理能提高数据质量，从而提高挖掘结果的质量。有大量数据预处理技术，主要有以下几种：

- 数据清理：数据清理可以去掉数据中的噪声，纠正不一致。
- 数据集成：数据集成将多个数据源中的数据结合起来存放在一个一致的数据存储中。
- 数据变换：数据变换将数据转换成适合于挖掘的形式。

其次是数据规范化(在一定程度上它也属于数据变换)。它有多种方法，例如：最小-最大规范化、z-score 规范化(或零—均值规范化)、小数定标规范化。

其实数据挖掘的步骤并不都是必须如此，根据实际情况可有灵活的变化，所以上面的步骤只是一般的过程。

第 3 章 AFS 理论基础知识

3.1 AFS 理论简介

模糊概念的描述、模糊逻辑控制器、模糊聚类分析、近似推理和其他人工智能等模糊集理论的应用领域都是建立在隶属函数和模糊逻辑基础上的，因而自 1965 年 L. A. Zadeh 提出模糊集理论以来，国内外许多学者在隶属函数和模糊逻辑方面做了大量的研究工作。然而，目前所采用的确定隶属函数的方法，主要都是靠人的经验和直觉，含有大量的人为因素。目前的模糊逻辑系统中的逻辑运算分别是用 T -模、 S -模及否定算子定义的，它们有无穷多种。逻辑运算算子的具体选取也含有一定的人为因素，由此导致了隶属函数和模糊逻辑系统的建立缺少客观性、严密性和统一性，使得许多重要的数学工具不能发挥其应有的作用，严重地阻碍和限制了模糊集与系统理论的发展和应用。

刘晓东教授于文献^[10-18]中提出的 AFS 理论，将人对概念的分解及合成的部分机理数学抽象化，较好地揭示了人靠经验和直觉描述复杂的模糊概念以及确定相应模糊概念的隶属函数的内在机理，不是用 T -模、 S -模和否定算子，而是由原始数据用 AFS 代数的运算和其上的一个逆序对合运算来建立模糊逻辑系统，同时已初步用拓扑分子格刻划人类概念之间的抽象关系，从而使得应用代数、拓扑、组合数学等严密的经典数学方法来研究模糊集理论成为可能，也使得隶属函数和模糊逻辑系统的建立更具客观性、严密性和统一性。

AFS 理论将模糊集的思想数学公理化，是一个更接近于实际的数学工具。AFS 理论给出了三个新的数学对象：AFS 代数、AFS 结构和认知域。AFS 代数是一种分子格，包含 EI 代数、 $'EI$ 代数、 EI'' 代数和 $'EI''$ 代数，它是概念的表示、结构、合成、分解的数学刻划，是建立 AFS 理论的数学基础。AFS 结构是论域 X 上的复杂关系的数学抽象，它和 AFS 代数结合可以给出论域 X 上的模糊集的隶属函数的表示以及 AFS 模糊逻辑系统的构造，AFS 结构与计算机的数据库在结构上很相似，非常便于计算机操作。而认知域则将 L -模糊集转化为 Zadeh-模糊集，即隶属度取值在 $[0, 1]$ 中的模糊集。这种 L -模糊集到 Zadeh-模糊集的转化，将使得模糊概念的 AFS 描述更直观，更易于操作。

在本章中, 我们主要从 EI 代数、 *EI 代数及其性质和 AFS 方法两个方面对 AFS 的理论进行了介绍。

3.2 EI 代数和 *EI 代数

AFS 代数是概念的表示、结构、合成、分解的数学刻画, 是建立 AFS 理论的数学基础, 它包含 EI 代数、 *EI 代数、 EI^n 代数和 $^*EI^n$ 代数。本节的主要内容, 就是介绍 EI 代数和 *EI 代数的性质、代数结构。先给出一些定义。本文只给出 EI 代数的基本概念, 对于 *EI 代数的相关概念可参见文献[10]。

设 M 是一非空集合, 定义 $EM = \{ \sum_{i \in I} A_i \mid A_i \in 2^M, i \in I, I \text{ 是任意指标集} \}$, 这里 $\sum_{i \in I} A_i$ 是形式和。如 I 是有限集, 即 $I = \{1, 2, \dots, n\}$, 则 $\sum_{i \in I} A_i$ 可记为 $A_1 + A_2 + \dots + A_n$, 即 $\sum_{i \in I} A_i = A_1 + A_2 + \dots + A_n$ 。当 $A_i (i \in I)$ 按照不同的次序作和时, $\sum_{i \in I} A_i$ 表示 EM 中相同的元素, 例如, $\sum_{i \in \{1,2\}} A_i = A_1 + A_2 = A_2 + A_1$, 在 EM 上定义二元关系 $R: \forall \sum_{i \in I} A_i, \sum_{j \in J} B_j \in EM, (\sum_{i \in I} A_i) R (\sum_{j \in J} B_j) \Leftrightarrow \forall A_i (i \in I), \exists B_h (h \in J)$ 使得 $A_i \supseteq B_h$, 并且 $\forall B_j (j \in J), \exists A_u (u \in I)$ 使得 $B_j \supseteq A_u$ 。

显然, R 是一个等价关系。不失一般性, 下面总用 EM 表示 EM/R , 此时, $\sum_{i \in I} A_i = \sum_{j \in J} B_j$ 意味着包含 $\sum_{i \in I} A_i$ 的等价类与包含 $\sum_{j \in J} B_j$ 的等价类相等。

命题 3.2.1 设 M 是一非空集合, $\sum_{i \in I} A_i \in EM$ 。若 $A_r \subseteq A_s, r, s \in I, r \neq s$, 则 $\sum_{i \in I} A_i = \sum_{i \in I, i \neq s} A_i$ 。

定义 3.2.1 设 M 是一非空集合, $\sum_{i \in I} A_i \in EM$ 。若 $\sum_{i \in I} A_i \neq \sum_{i \in I, i \neq u} A_i \quad \forall u \in I$, 则称 $\sum_{i \in I} A_i$ 是既约的。

命题 3.2.2 设 M 是一非空集合, $\sum_{i \in I} A_i, \sum_{j \in J} B_j \in EM$ 。若 $\sum_{i \in I} A_i$ 和 $\sum_{j \in J} B_j$ 都是既约的, 并且 $\sum_{i \in I} A_i = \sum_{j \in J} B_j$, 则 $\{B_j \mid j \in J\} = \{A_i \mid i \in I\}$ 。

定义 3.2.2 设 M 是一非空集合。 $\forall \sum_{i \in I} A_i, \sum_{j \in J} B_j \in EM$, 定义

$$\sum_{i \in I} A_i * \sum_{j \in J} B_j = \sum_{i \in I, j \in J} A_i \cup B_j,$$

$$\sum_{i \in I} A_i + \sum_{j \in J} B_j = \sum_{k \in I \cup J} C_k,$$

这里 $I \cup J$ 是 I 和 J 的不交并. 若 $k \in I$, 则 $C_k = A_k$, 若 $k \in J$, 则 $C_k = B_k$.

设 $\sum_{i \in I} A_i \in EM$, $j = 1, 2, \dots, n$, 以后我们总是记:

$$* \prod_{j=1}^n (\sum_{i \in I_j} A_i) = (\sum_{i \in I_1} A_i) * (\sum_{i \in I_2} A_i) * \dots * (\sum_{i \in I_n} A_i).$$

EM 上的 “*” 和 “+” 运算满足如下性质: $\forall \sum_{i \in I} A_i \in EM$,

(1) $\emptyset * (\sum_{i \in I} A_i) = \sum_{i \in I} A_i, \forall \sum_{i \in I} A_i \in EM;$

(2) $M * (\sum_{i \in I} A_i) = M, \forall \sum_{i \in I} A_i \in EM;$

(3) $\emptyset + \sum_{i \in I} A_i = \emptyset, \forall \sum_{i \in I} A_i \in EM;$

(4) $M + \sum_{i \in I} A_i = \sum_{i \in I} A_i, \forall \sum_{i \in I} A_i \in EM.$

定义 3.2.3 设 M 是一非空集合. $\forall \sum_{i \in I} A_i, \sum_{j \in J} B_j \in EM$, 定义:

$$(\sum_{i \in I} A_i) \succeq (\sum_{j \in J} B_j) \text{ (或 } (\sum_{j \in J} B_j) \preceq (\sum_{i \in I} A_i)) \Leftrightarrow \forall B_j (j \in J), \exists A_u (u \in I) \text{ 使得 } B_j \supseteq A_u.$$

A_u .

可以验证, “ \succeq ” 是 EM 上的偏序关系。

我们称 (EM, \succeq) 为集 M 上的 EI 代数。

命题 3.2.3 设 M 是一非空集合. $\forall \sum_{i \in I} A_i, \sum_{j \in J} B_j \in EM$, 则

$$(\sum_{i \in I} A_i) \succeq (\sum_{j \in J} B_j) \Leftrightarrow \sum_{i \in I} A_i + \sum_{j \in J} B_j = \sum_{i \in I} A_i \Leftrightarrow (\sum_{i \in I} A_i) * (\sum_{j \in J} B_j) = \sum_{j \in J} B_j.$$

定理 3.2.1 设 M 是一非空集合, 则 (EM, \succeq) 是一有泛界的分配格。

其中 $\sum_{i \in I} A_i + \sum_{j \in J} B_j$ 是 $\sum_{i \in I} A_i$ 与 $\sum_{j \in J} B_j$ 的上确界, $\sum_{i \in I} A_i * \sum_{j \in J} B_j$ 是 $\sum_{i \in I} A_i$ 与 $\sum_{j \in J} B_j$ 的下

确界, \emptyset 是最大元, M 是最小元。

定理 3.2.2 设 M 是一非空集合, 则 (EM, \succeq) 是分子格, 且 $\forall A \in 2^M, A$ 是 + 一既约元 (v —既约元)。

定理 3.2.3 设 M 是一非空集合且 $|M| > 1$, 则 (EM, \geq) 不是 Boole 代数。

定义 3.2.5 设 L 是一完备格, $a \in L, B \subseteq L$ 。如果 $B \neq \emptyset$ 并且

(1) $\sup B = a$;

(2) $\forall A \subseteq L$, 若 $\sup A \geq a$, 则 $\forall x \in B, \exists y \in A$ 使得 $y \geq x$;

则称 B 为 a 的一个极小族。

定义 3.2.6 设 L 是一完备格, $a \in L, B \subseteq L$ 。如果 B 是 a 的一个极小族且 B 的元素都是 v -既约元, 则称 B 是 a 的标准极小族。

3.3 AFS 方法的基本思想和模糊概念的 AFS 表示

3.3.1 基本概念

AFS 方法的基本思想是: 用极少的简单概念或属性, 按 AFS 代数的运算法则表示非常多的复杂概念。本节首先通过一个实例概括性地介绍和分析 AFS 方法的基本思想、数学本质和抽象定义的工程意义, 然后讨论模糊概念的 AFS 表示问题。

定义 3.3.1 设 X, M 为两个集合, 2^M 是 M 的幂集, $\tau: X \times X \rightarrow 2^M$ 。如果 τ 满足

$$AX1: \forall (x_1, x_2) \in X \times X, \tau(x_1, x_2) \subseteq \tau(x_1, x_1)$$

$$AX2: \forall (x_1, x_2), (x_2, x_3) \in X \times X, \tau(x_1, x_2) \cap \tau(x_2, x_3) \subseteq \tau(x_1, x_3)$$

则 (M, τ, X) 被称为一个 AFS 结构, 称 X 为论域, 称 M 为属性集, 称 τ 为结构。

定义 3.3.2 设 R 为 X 上的二元关系, 如果 R 满足:

(1) $\forall x \in X, (x, x) \in R$;

(2) 若 $(x, y), (y, z) \in R$, 则 $(x, z) \in R$;

(3) $\forall x, y \in X$, 或者 $(x, y) \in R$, 或者 $(y, x) \in R$;

则称 R 为 X 上的偏好关系。

Kim K. H. 在文献[19]中对偏好关系进行了系统深入的研究, 并且指出许多概念与偏好关系相对应。每个偏好关系 R 可以用如下的强度链来表示:

$$\dots u > v > x = y > z \dots$$

其中 $x = y$ 并不表示 x, y 为 X 中的同一元素, 而是表示 $(x, y) \in R$ 且 $(y, x) \in R$,

即对 x, y 的强度相等。

定义 3.3.3 设 X 和 M 为两个集合, (M, τ, X) 是一个 AFS 结构, (X, σ, m) 是一个正的测度空间, 即 $0 < m(X) < +\infty$, 则 (M, τ, X, σ, m) 被称为一个半认知域。如果 $a = \sum_{i \in I} A_i \in EM$, 且对 $\forall x \in X, i \in I$, 有 $\underline{A}_i(\{x\}) \in \sigma$, 则称概念 a 在该半认知域上可测, 并且其隶属函数为:

$$\mu_a(x) = \sup_{i \in I} (m(\underline{A}_i(\{x\})) / m(X)), \forall x \in X$$

文献[11]用 AFS 代数的拓扑及拓扑分子格结构研究分析了模糊概念之间的关系, 并且将文献[9]中给出的拓扑分子格理论与复杂的实际问题联系起来。下面是 L -模糊集隶属函数的表示定理。

定理 3.3.1 设 (M, τ, X) 是一个 AFS 结构, $B \subseteq X, A \subseteq M$,

$$\underline{A}(B) = \{y \mid y \in X, \tau(x, y) \supseteq A, \forall x \in B\}, x \in X.$$

对 $\forall \sum_{i \in I} A_i \in EM$, 定义 $\phi_x(\sum_{i \in I} A_i) = \sum_{i \in I} \underline{A}_i(\{x\}) A_i \in EXM$, 则 ϕ_x 是从 (EM, \wedge, \vee) 到 (EXM, \wedge, \vee) 上的代数同构。

因为 (EXM, \wedge, \vee) 是一个格, 所以由前面的定理可知, $\forall \sum_{i \in I} A_i \in EM$ 都是论域 X 上的一个 L -模糊集。 $\forall x \in X, x$ 属于模糊集 $\sum_{i \in I} A_i$ 的格值隶属度为 $\sum_{i \in I} \underline{A}_i(\{x\}) A_i \in EXM$. 对于每一个模糊集 $\zeta \in EM$ 及 $x, y \in X$, 如果 $\zeta(x) \wedge \zeta(y) = \zeta(y)$, 即在分子格 (EXM, \wedge, \vee) 中 $\zeta(x) \geq \zeta(y)$, 则 x 属于模糊集 ζ 的程度大于或等于 y 属于模糊集 ζ 的程度, “ \wedge ”和“ \vee ”分别为 EM 内模糊集的逻辑算子“and”和“or”。

3.3.2 AFS 方法的基本思想及工程意义

例 3.1 设 $X = \{x_1, x_2, \dots, x_{10}\}$ 是 10 个人的集合, $M = \{m_1, m_2, \dots, m_{11}\}$, 其中 $m_1 =$ 老, $m_2 =$ 身高高, $m_3 =$ 体重重, $m_4 =$ 工资高, $m_5 =$ 财富多, $m_6 =$ 男性, $m_7 =$ 女性, $m_8 =$ 肤色黑, $m_9 =$ 肤色白, $m_{10} =$ 肤色黄, $m_{11} =$ 信用高. 关于论域 X 和属性集 M , 下面有关于肤色黑、白、黄的偏好关系和表 3.1 的属性关系:

肤色黑的程度由强到弱: $x_7 > x_{10} > x_4 = x_8 > x_2 = x_9 > x_5 > x_6 = x_3 = x_1$;

肤色白的程度由强到弱: $x_6 = x_3 = x_1 > x_5 > x_2 = x_9 > x_4 = x_8 > x_{10} > x_7$;

肤色黄的程度由强到弱： $x_2 = x_9 > x_4 = x_8 = x_5 > x_{10} > x_6 = x_3 = x_1 = x_7$ 。

考察例 3.1 中给出的各个属性可以看出，像“工资”、“财富”、“肤色”这样的概念简单且满足偏好关系。但像“信用高”这样的模糊概念则不然，它既复杂又可能不满足偏好关系(即有时是不可比的)。那么，如何构造一个模糊集来表达这类模糊概念呢？下面我们以“信用高”为例来说明。

“信用高”的评价有许多标准，例如一个高工资年龄大的男性，或一个财富多年龄大的男性，或一个高工资的女性，或一个财富多的女性等等。当然不同的人会有不同的描述，为了简明起见，在本例中不妨假设：信用高 = 高工资大龄的男性，或财产多年龄大的男性，或高工资的女性，或财产多的女性。因为 x_6 和 x_9 都是女性且 x_6 比 x_9 工资高，由假设知 x_6 比 x_9 信用高； x_9 和 x_3 都是女性且 x_9 比 x_3 财产多，由假设知 x_9 比 x_3 信用高；但由表 3-1 知道 x_6 不比 x_3 信用高，所以“信用高”不满足偏好关系定义中的(2)，即“信用高”不能像肤色等概念那样用一个偏好关系来描述。

表 3.1 属性关系

Tab. 3.1 relations of factors

	年龄	身高	体重	工资	财产	男性	女性
x_1	20	1.9	90	1	0	是	否
x_2	13	1.2	32	0	0	否	是
x_3	50	1.7	67	140	34	否	是
x_4	80	1.8	73	20	80	是	否
x_5	34	1.4	54	15	2	是	否
x_6	37	1.6	80	80	28	否	是
x_7	45	1.7	78	268	90	是	否
x_8	70	1.65	70	30	45	是	否
x_9	60	1.82	83	25	98	否	是
x_{10}	3	1.1	21	0	0	否	是

根据表 3.1 和肤色的偏好关系描述, 按如下原则构造 AFS 结构 $(M \setminus \{m_{11}\}, \tau, X)$: $\forall x, y \in X, m \in M \setminus \{m_{11}\}, m \in \tau(x, y) \Leftrightarrow x$ 具有性质 m 且 x 具有性质 m 的程度强于或等于 y 具有性质 m 的程度. 在这里限制 $m \in M \setminus \{m_{11}\}$, 是因为若 $m = m_{11}$, 则由表 3-1 有 $m_{11} \in \tau(x_6, x_9) \cap \tau(x_9, x_3)$, 但 $m_{11} \notin \tau(x_6, x_3)$, 即 τ 不满足定义 3.3.1 中的 AX2. 容易验证, $\forall m \in M \setminus \{m_{11}\}$, 则满足定义 3.3.1 的 AX1, AX2, 即 τ 满足定义 3.3.1.

AFS 结构 (M, τ, X) 是论域 X 的复杂关系的数学抽象, 包含在表 3.1 中和关于肤色的偏好关系中几乎所有的信息都转化到 (M, τ, X) 中, 用它和文献[14]中给出的 AFS 代数可以给出论域 X 上的模糊集的隶属函数的表示以及构造 AFS 模糊逻辑系统. 许多实际问题都是用表 3.1 那样的数据库表示的.

根据我们建立的例 3.1 中的 AFS 结构 (M, τ, X) , 对模糊集“信用高(credit)” $\text{credit} = \{m_1, m_4, m_6\} + \{m_1, m_5, m_6\} + \{m_4, m_7\} + \{m_5, m_7\} \in EM$ 的格值隶属度进行计算, 得

$$\begin{aligned} \text{credit}(x_1) &= \{x_1, x_2, x_{10}\} \{m_1, m_4, m_6\} + \emptyset \{m_1, m_5, m_6\} + \emptyset \{m_4, m_7\} + \emptyset \{m_5, m_7\} \\ \text{credit}(x_2) &= \emptyset \{m_1, m_4, m_6\} + \emptyset \{m_1, m_5, m_6\} + \emptyset \{m_4, m_7\} + \emptyset \{m_5, m_7\} \\ \text{credit}(x_i) \quad (i=3, 4, \dots, 10) &\text{ 略} \end{aligned}$$

通过分子格 EXM , 有时可以比较两点的隶属度的大小, 例如由 $\text{credit}(x_5) \wedge \text{credit}(x_8) = \text{credit}(x_5)$, 知 x_8 属于 credit 的程度大于 x_5 ; 但有时也不能比较两点的隶属度的大小, 例如, 因为 $\text{credit}(x_4) \wedge \text{credit}(x_3)$ 不等于 $\text{credit}(x_4)$ 和 $\text{credit}(x_3)$ 中的任何一个, 所以 x_3 和 x_4 属于 credit 的程度在分子格 EXM 中不能比较. 但总的来说, X 中的元素按属于 credit 程度的大小关系可以构成一个有向图.

为了更直观, 进一步地还可以用前面定义的半认知域 L —模糊集转化为 Zadeh—模糊集, 即隶属度取值在 $[0, 1]$ 中的模糊集.

事实上, 对 $a, b \in EM$, 如果在分子格 EM 中 $a \geq b$, 则在分子格 EXM 中 $a(x) \geq b(x)$ 且 $\mu_a(x) \geq \mu_b(x), x, y \in X$;

如果在分子格 EXM 中 $a(x) = b(y)$, 则: $\mu_a(x) = \mu_b(y), x, y \in X$;

如果对 $\forall x \in X, \mu_a(x) = \chi_A(x), \mu_b(x) = \chi_B(x)$, 其中 $A, B \subseteq X$, 则:

$$\mu_{a \wedge b}(x) = \chi_{A \cap B}(x), \mu_{a \vee b}(x) = \chi_{A \cup B}(x), \forall x \in X$$

格值隶属函数、逻辑运算 \wedge, \vee 都与人类的直觉非常接近. 我们已经证明^[14]:

EM 包含至少 $\sum_{i=1}^n (2^{c_i} - 1)$ 个模糊集, 即当属性集 M 包含 n 个元素时, 用 n 个简单的属性可以表示至少 $\sum_{i=1}^n (2^{c_i} - 1)$ 个复杂的模糊概念。比如用例 3.1 中的 10 个简单属性 $\{m_1, m_3, m_4, \dots, m_{11}\}$, 可以表示至少 10^{37} 像“信誉高”那样的复杂模糊概念。

第 4 章 AFS 模糊逻辑聚类分析方法

文[51]提出了 AFS 模糊逻辑聚类分析算法,它在文[44]的基础上进行了进一步的研究,对其算法进行了改进。本章我们来介绍一下文[51]提出的 AFS 模糊逻辑聚类分析算法。

4.1 AFS 模糊逻辑聚类分析算法

X 是对象集, M 是属性集, (M, τ, X) 是由原始数据建立 AFS 结构,关于如何由原始数据得到 AFS 结构,请参考第 3 章中的例子。 $\Lambda \subseteq EM$ 。下面给出改进后的算法:

Step1:

$\forall x \in X$, $\mu_{\theta}(x)$ 是 x 在聚类分析中属于某一类的最大隶属度,其中 $\theta = \bigvee_{b \in A} b$ 。

Step 2:

对于每个 $x \in X$, 求出 x 的模糊描述 ζ_x 。 ζ_x 满足: $\zeta_x \in (\Lambda)_{EI}$, 其中 $(\Lambda)_{EI}$ 是 Λ 用 AFS 模糊逻辑生成的 EM 的 EI 子代数(定义 2.3): $\mu_{\zeta_x}(x) \geq \mu_{\theta}(x) - \varepsilon$, ε 是一个充分小的正数或零,并且对于 $y \in X$, $y \neq x$, $\mu_{\zeta_x}(y)$ 尽可能地小。换句话说,用 $(\Lambda)_{EI}$ 中的模糊集把 x 与 X 中其他元素最大限度地区分开。那些无法与 x 区分的元素为 x 的同类。

下面给出求点 x 的模糊描述的算法步骤(当 M 为有限集):

1. 因为 M 有限且 $\Lambda \subseteq EM$, 所以可设 $\Lambda = \{a_i \mid a_i = \sum_{k=1}^{n_i} A_k^i \in EM\}$, 将 Λ 中所有模糊集都化简为不可约的。

2. 选取适当的 $\varepsilon \geq 0$, ε 应充分小。当 Λ 有限时可令 $\varepsilon = 0$ 。对任意 $a_i = \sum_{k=1}^{n_i} A_k^i \in \Lambda$, 选出 $\mu_{A_l^i}(x) \geq \mu_{\bigvee_{k=1}^{n_i} A_k^i}(x) - \varepsilon, 1 \leq l \leq n_i$, 将 A_l^i 放入一个集合,记为 B_x^{ε} 。

3. 选取极大的集合 $H \subseteq B_x^{\varepsilon}$, 使得 $\mu_{\bigwedge_{\beta \in H} \beta}(x) \geq \mu_{\bigvee_{k=1}^{n_i} A_k^i}(x) - \varepsilon$, 然后将分子 $\bigwedge_{\beta \in H} \beta$ 放入一个集合,记为 A_l^i 。

4. 最后得到 x 在模糊特征 Λ 下的模糊描述 ζ_x 。在 Λ^c ($\Lambda^c = \{A_1, A_2, \dots, A_m\}$) 中选取能够描述 x 的最好的分子作为 x 在模糊特征 Λ 下的模糊描述。具体的说, 求出所有样本中属于 A_i ($i=1, 2, \dots, m$) 的隶属程度大于 δ ($\delta \in (0, 1)$) 的样本总数, 记为 n_i ($i=1, 2, \dots, m$)。从 n_1, n_2, \dots, n_m 中找到最小的数 n_k , 那么 n_k 所对应的 A_k 即为最后 x 在模糊特征 Λ 下的模糊描述, 记为 ζ_x 。(在实例中, 令 $\delta = 0.5$ 。)

Step 3:

根据 x 的模糊描述 ζ_x 给出任意 $x, y \in X$, x 与 y 相关的程度

$$m_{xy} = \min\{\mu_{\zeta_x, \zeta_{x_j}}(x_j), \mu_{\zeta_{x_j}, \zeta_x}(x_j)\}$$

其中 $X = \{x_1, x_2, \dots, x_n\}$, $M_\Lambda = (m_{ij})$ 为 X 上模糊关系矩阵。能够证明存在整数 k 使得 $(M_\Lambda^k)^2 = M_\Lambda^k$ 。这保证了由模糊关系矩阵 $Q = M_\Lambda^k$ 可以导出论域 X 上的等价关系。

Step 4:

$Q = M_\Lambda^k = (q_{ij})$, Boolean 矩阵 $Q_\alpha = (q_{ij}^\alpha)$, 其中 $q_{ij}^\alpha = 1 \Leftrightarrow q_{ij} \geq \alpha$, $\alpha \in [0, 1]$ 。 $x_i, x_j \in X$, x_i 与 x_j 以置信度 α 等价 (即为同一类) 当且仅当 $q_{ij}^\alpha = 1$; 若 $q_{ij}^\alpha = 0$, 则对于给定的置信度 α , 不能判断 x_i 所属的类。

Step 5:

对于在置信度 α 下 X 被分为 C_1, C_2, \dots, C_l 类,

$$A_{C_i}^\alpha = \bigcup_{x \in C_i} \{A \mid A \in \Lambda^c, \mu_A(x) \geq \alpha\}$$

$$\zeta_{C_i} = \bigvee_{A \in A_{C_i}^\alpha} A$$

模糊集 ζ_{C_i} 为类 C_i 的模糊描述。求出所有样本属于每类模糊描述的隶属度, 进行比较。每个样本属于哪类模糊描述的隶属度最大, 那么该样本就属于哪类。这样可以进一步对样本聚类。模糊集 ζ_{bou} 描述了模糊分类的所有类之间的边界, 其中

$$\zeta_{bou} = \bigvee_{1 \leq i, j \leq l, i \neq j} (\zeta_{C_i} \wedge \zeta_{C_j})$$

指标 I_α 可用于评价以 α 为置信度分类结果的好坏, 其中

$$I_{\alpha} = \frac{\sum_{x \in \bigcup_{i=1}^{l_{\alpha}} C_i} \mu_{\zeta_{i_{\alpha}}}^{\alpha}(x)}{\sum_{x \in \bigcup_{i=1}^{l_{\alpha}} C_i} \mu_{\zeta_{i_{\alpha}}}^{\alpha}(x)}$$

其中 $\zeta_{Total} = v_{1_{\alpha}} \zeta_{C_i}$, $i \geq 2$ 。当 I_{α} 最小时, 分类最清晰。

4.2 算法对 iris (杉树) 数据进行模糊聚类分析

Iris 的原始数据是 150 个样本, 4 个属性构成的数据集, 可用 150×4 的矩阵表示。矩阵的第一列为所有样本的萼片长度, 第二列为所有样本的萼片宽度, 第三列为所有样本的花瓣长度, 第四列为所有样本的花瓣宽度。样本的原始数据按产地分为三类: x_1, x_2, \dots, x_{50} 属于 Iris Setosa, $x_{51}, x_{52}, \dots, x_{100}$ 属于 Iris Versicolour, $x_{101}, x_{102}, \dots, x_{150}$ 属于 Iris Virginica。

下面建立该例的 AFS 结构 (M, τ, X)

(1) $X = \{x_1, x_2, \dots, x_{150}\}$ 是 150 个样本的集合。

(2) $M = \{m_1, m_2, \dots, m_8\}$, 其中 m_1 为花瓣长度大即花瓣长, m_2 为花瓣长度不大即花瓣短, m_3 为花瓣长度适中, m_4 为花瓣长度不适中, m_5 为花瓣宽, m_6 为花瓣窄, m_7 为花瓣宽度适中, m_8 为花瓣宽度不适中。

按下列规则构造 $\tau: X \times X \rightarrow 2^M$, $m \in M, x_i, x_j \in X$,

$m \in \tau(x_i, x_j) \Leftrightarrow x_i$ 属于 m 的程度大于或等于 x_j 属于 m 的程度。例: $m_3 \in \tau(x_i, x_j) \Leftrightarrow$

样本 x_i 的花瓣长度比样本 x_j 的花瓣长度更接近于 150 个样本花瓣长度的平均值或它们接近于 150 个样本花瓣长度平均值的程度相等。

(3) 令 $\sigma = 2^X$, $m(A) = |A| / (|A|)$ 等于集合 A 的元素个数, $A \in 2^X$, 则 (M, τ, X, σ, m) 为半认知域并给出所有 Zadeh 模糊集的隶属函数。

注:

1. 在建立 Iris 数据的 AFS 结构时, τ 可由属性上的序关系确定, 即只要所有属性能用序关系表示就可建立 AFS 结构。因此, 没有采用样本在属性上的数值而

是采用样本在属性上由数值导出的序关系。

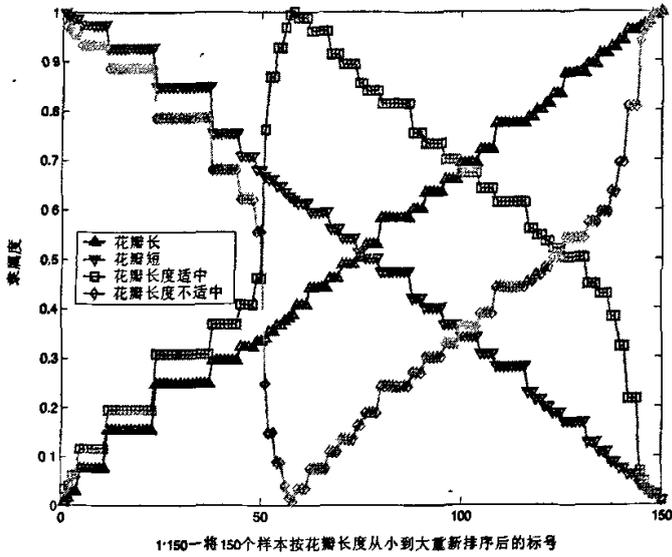


图 4.1 花瓣长度隶属度图形

Fig.4.1 Membership degree of the petal length

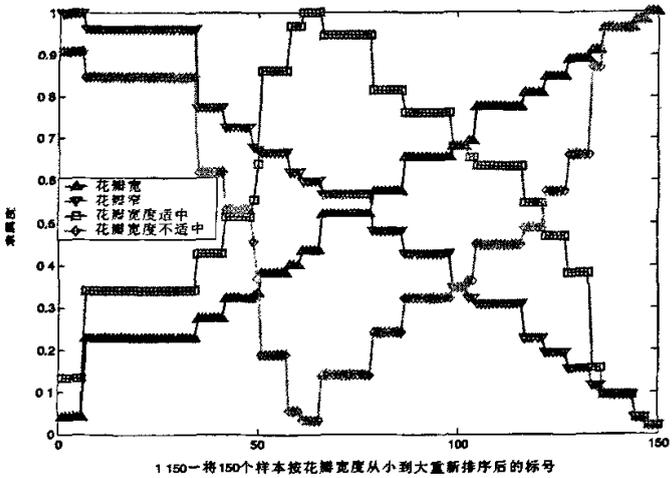


图 4.2 花瓣宽窄隶属度图形

Fig.4.2 Membership degree of the petal width

2. 将 150 个样本按花瓣长度从小到大排序，得到在此新顺序下的 1 到 150 个

样本，下图即为重新排序后的所有样本在属性“花瓣长”、“花瓣短”、“花瓣长度适中”和“花瓣长度不适中”上的隶属度图形。同理，再将 150 个样本按花瓣宽度从小到大排序，得到所有样本在属性“花瓣宽”、“花瓣窄”、“花瓣宽度适中”和“花瓣宽度不适中”上的隶属度图形，即图 4.1。

图 4.2 是所有样本在属性“花瓣宽”、“花瓣窄”、“花瓣宽度适中”和“花瓣宽度不适中”上的隶属度图形。

150 个样本在模糊特征下的模糊描述即：

$$A = \{m_1, m_2, m_3, m_4, m_5, m_6, m_7, m_8\} \subseteq EM$$

Step1:

求出模糊集 θ 获得每个对象属于某类的最大隶属度

$$\theta = \nu_{b \in A} b = \{m_1\} + \{m_2\} + \{m_3\} + \{m_4\} + \{m_5\} + \{m_6\} + \{m_7\} + \{m_8\}。$$

Step2:

取 $\varepsilon = 0.1$ ，有 $B_{x_1}^{0.1} = \{\{m_2\}, \{m_4\}, \{m_6\}\}$ ，

$$A_{x_1}^{0.1} = \{\{m_2, m_4\}, \{m_2, m_6\}\}。$$

根据算法 Step2(4)可知 x_1 在 A 下的模糊描述为 $\zeta_{x_1} = \{m_2, m_4\}$ 。下面列几个：

$$A_{x_{13}}^{0.1} = \{\{m_2, m_6\}, \{m_6, m_8\}\}, \quad \zeta_{x_{13}} = \{m_6, m_8\}。$$

$$A_{x_{53}}^{0.1} = \{\{m_1\}, \{m_3, m_7\}\}, \quad \zeta_{x_{53}} = \{m_3, m_7\}。$$

$$A_{x_{57}}^{0.1} = \{\{m_1\}, \{m_5\}, \{m_3, m_7\}\}, \quad \zeta_{x_{57}} = \{m_3, m_7\}。$$

$$A_{x_{124}}^{0.1} = \{\{m_3\}, \{m_1, m_5\}\}, \quad \zeta_{x_{124}} = \{m_1, m_5\}。$$

$$A_{x_{133}}^{0.1} = \{\{m_1, m_5\}, \{m_5, m_8\}\}, \quad \zeta_{x_{133}} = \{m_5, m_8\}。$$

Step3:

得模糊关系矩阵 $M_A = (r_{ij})$

$$M_A =$$

0.88667	0.88667	0.85333	...	0.42000	0.42000	0.42000	
	0.88667	0.85333	...	0.42000	0.42000	0.42000	
		0.90000	...	0.42000	0.42000	0.42000	
			⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
				0.72667	0.63333	⋮	
					⋮	0.70000	...	0.63333	⋮	
							⋮	0.63333	⋮	
								⋮	⋮	⋮	⋮	
									1	0.96000	...	0.72000
										0.96000	...	0.72000
											⋮	⋮
												0.72000

经验证 $M_A^2 = M_A$ 。

Step4—5:

用模糊矩阵给出所有样本按所选聚类属性的聚类分析。当 $\alpha = 0.64$ 即置信度为 0.64 时 Q_α 为 X 上的一个等价关系，它把 X 分为 3 类，第一类 C_1 的样本为： $x_1, x_2, \dots, x_{23}, x_{25}, x_{26}, \dots, x_{50}$ 。这类模糊描述的模糊集 ξ_{C_1} 为 $\{m_6\} + \{m_2, m_4\}$ ，即第一类为“花瓣窄的杉树”或“花瓣短且不适中的杉树”。第二类 C_2 的样本为： $x_{51}, x_{52}, \dots, x_{77}, x_{79}, x_{80}, \dots, x_{83}, x_{85}, x_{86}, \dots, x_{100}, x_{107}, x_{127}, x_{139}$ 。这类模糊描述的模糊集 ξ_{C_2} 为 $\{m_7\} + \{m_3\}$ ，即第二类为“花瓣宽度适中的杉树”或“花瓣长度适中的杉树”。第三类 C_3 的样本为： $x_{78}, x_{84}, x_{101}, x_{102}, \dots, x_{106}, x_{108}, x_{109}, \dots, x_{126}, x_{128}, x_{129}, \dots, x_{138}, x_{140}, x_{141}, \dots, x_{150}$ 。这类模糊描述的模糊集 ξ_{C_3} 为 $\{m_1\} + \{m_5\}$ ，即第三类为“花瓣长的杉树”或“花瓣宽的杉树”。

求出所有样本分别属于这三类模糊描述的隶属度(见图 4.3)，进行比较，可以得到新的聚类结果。会发现上次聚类丢掉的 x_{24} 现在属于 C_1 类，原来属于 C_2 类的 x_{71}, x_{127}, x_{139} ，现在属于 C_3 类，原来属于 C_3 类的 x_{120} ，现在属于 C_2 类。最终分类结果的出错样本分别为 $x_{71}, x_{78}, x_{84}, x_{107}, x_{120}$ ，正确率达到 96.67%。

$$\zeta_{\text{base}} = \{m_2, m_4, m_7\} + \{m_6, m_7\} + \{m_2, m_3, m_4\} + \{m_3, m_6\} + \{m_2, m_4, m_5\} + \{m_5, m_6\} + \{m_1, m_2, m_4\} + \{m_1, m_6\} + \{m_5, m_7\} + \{m_3, m_5\} + \{m_3, m_7\} + \{m_1, m_3\},$$

$$\zeta_{\text{Total}} = \{m_2, m_4\} + \{m_6\} + \{m_7\} + \{m_3\} + \{m_5\} + \{m_1\},$$

$$I_\alpha = 0.4823.$$

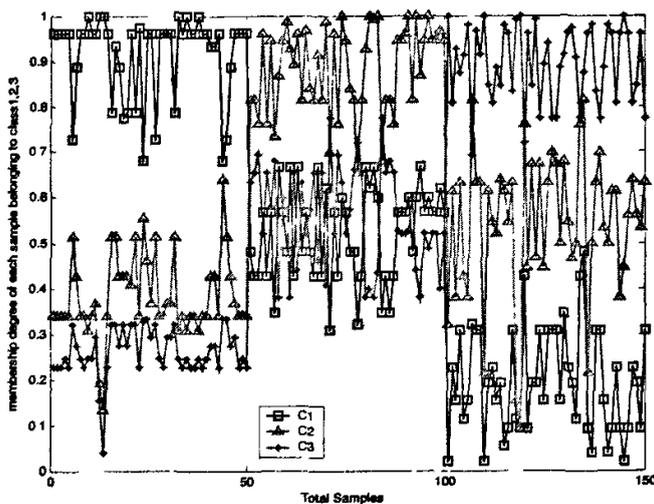


图 4.3 所有样本在置信度 α 为 0.64 时得到的三类模糊描述下的隶属度图形

Fig.4.3 Membership degree of the three of fuzzy class when $\alpha = 0.64$

当 $\alpha = 0.43$ 时，由 Q_α 得到 X 上的 2 聚类：第一类 C_1 的样本为： x_1, x_2, \dots, x_{50} 。这类模糊描述的模糊集 ζ_{C_1} 为 $\{m_6\} + \{m_2, m_4\}$ ，即一类为“花瓣窄的杉树”或“花瓣短且不适宜的杉树”。第二类 C_2 的样本为： $x_{51}, x_{52}, \dots, x_{150}$ 。这类模糊描述的模糊集 ζ_{C_2} 为 $\{m_7\} + \{m_3\} + \{m_1\} + \{m_5\}$ ，即 C_2 类为“花瓣宽度适中的杉树”或“花瓣长度适中的杉树”或“花瓣长的杉树”或“花瓣宽的杉树”。

求出所有样本分别属于这两类的隶属度(见图 4.4)，进行比较，聚类结果没有改变。

$$\xi_{\text{bow}} = \{m_2, m_4, m_7\} + \{m_6, m_7\} + \{m_2, m_3, m_4\} + \{m_3, m_6\} + \{m_2, m_4, m_5\} + \{m_3, m_6\} + \{m_1, m_2, m_4\} + \{m_1, m_6\} + \{m_5, m_7\} + \{m_3, m_5\} + \{m_1, m_7\} + \{m_1, m_3\},$$

$$\xi_{\text{Total}} = \{m_2\} + \{m_6\} + \{m_7\} + \{m_3\} + \{m_5\} + \{m_8\} + \{m_1\},$$

$$I_\alpha = 0.3750.$$

当 $\alpha = 0.73$ 时, $I_\alpha = 0.6196$ 。

当 $\alpha = 0.96$ 时, $I_\alpha = 0.6828$ 。

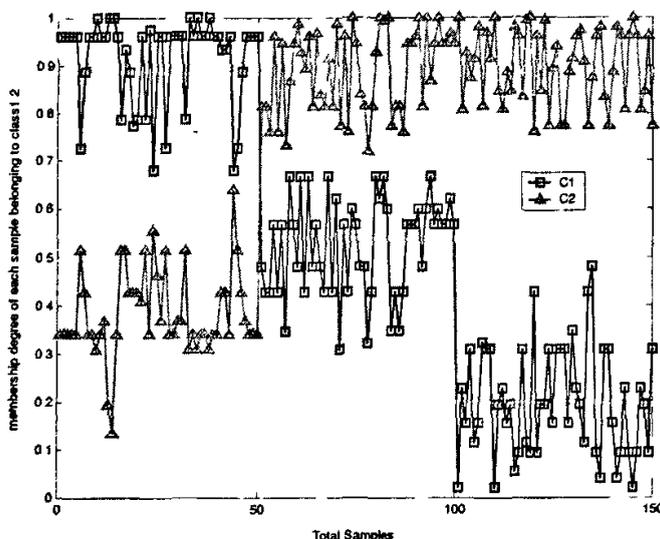


图 4.4 所有样本在置信度 α 为 0.43 时得到的两类模糊描述下的隶属度图形

Fig.4.4 Membership degree of the two of fuzzy class when $\alpha = 0.43$

从指标 I_α 可判断以置信度 $\alpha = 0.43$ 分类的结果最清晰。

观察图 c 可以得到以下结论：

(1) 当置信度 α 取 0.64, 即当把 150 个样本分为三类时得到的结果与 Iris 样本按产地分得到的数据分布结果 (见图 4.5) 基本上是一致的。得到的结果是第一类全部正确, 第二类有三个样本被分到第三类里面, 第三类有两个样本被分到第二类, 与图 4.5 显示的结果基本相同。

(2) 观察图 4.5 发现, 当把 150 个样本分为两类时, 结果最清晰。这与我们通过判断指标 I_a 得到的以置信度 $\alpha = 0.43$ 分类的结果最清晰的结论完全一致, 因此也说明指标 I_a 是非常准确的。

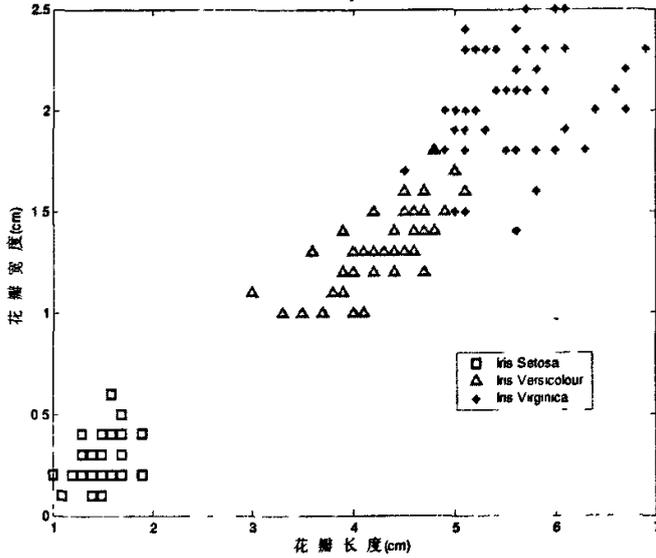


图 4.5 所有样本在花瓣长度和花瓣宽度下的图像

Fig.4.5 Membership degree of the petal length and width

第 5 章 AFS 模糊逻辑分类分析方法

在现今的模糊理论中,隶属函数通常是通过人的直觉获得,通过 t 模和 t 余模来定义模糊逻辑运算的。但是在实际应用中,大规模的智能系统往往是非常大和复杂的,包含了大量的概念,而这些概念不可能通过人的直觉来定义其隶属函数。刘晓东教授在文[44]中,基于 AFS 理论,提出了确定隶属度以及它们的逻辑运算的一种新的算法。在 AFS 理论中,隶属函数以及它们的逻辑运算都是通过 AFS 结构和 AFS 代数来确定的。文[44]用简单的例子揭示和强调了这种算法的优点。

本文在确定隶属度时,只用到了样本属性上的序关系,而没有用到样本属性上的数值。这种方法的好处在于,对于那些无法给出数值的概念,我们同样能够确定其隶属度,如几个人头发黑的程度,我们可以给出它们的序关系,而无法准确的给出它们的数值。本文中的例子能够证明这种用序关系确定隶属度的方法在实际应用中是非常有效的,应用的范围也比用数据确定隶属度的范围要广。

文[51]中提出了一种基于 AFS 模糊逻辑的聚类分析算法,并且将改进后的算法应用到样本数目较大的 Iris 数据中,得到了较好的分类结果。本文在文[51]的基础上进行研究。

文[44]中,作者在 AFS 的理论框架内提出了 AFS 模糊逻辑聚类分析算法,并给出了置信度 α 的指标:

$$I_{\alpha} = \frac{\sum_{x \in \bigcup_{i=1}^m C_i} \mu_{t_{\text{down}}}^{\alpha}(x)}{\sum_{x \in \bigcup_{i=1}^m C_i} \mu_{t_{\text{total}}}^{\alpha}(x)},$$

由于这一指标不是对每个样本都做考虑,将对这一指标进行完善。同时在第四章提到的 AFS 模糊逻辑聚类分析算法的基础上,应用 Iris 数据对 α 的选取做了进一步研究。同时应用改进后的 AFS 模糊逻辑聚类分析算法来设计一种新的模糊分类器,所设计的基于 AFS 模糊逻辑的模糊分类器不仅在本质上与以往的方法不同,而且可以将其应用到那些样本属性只能用序关系描述而无法用数值描述的数据集。因为以往的分类器大多只是适用于样本属性用数值来描述的数据集,所以对于样本属性用序关系描述的就非常困难或根本不可能处理。分类器设计过程的每一步都与人的认识过程类似,由于分类器的表示形式具有明确的语义,所以完

全可被每个人所理解。

并将这种新的模糊分类器应用于 Iris 数据。在 Iris 数据中随机选取 60%数据来设计分类器，余下的 40%数据作为测试样本，准确率达到了 98%。最后，我们又利用 Iris 数据对这种新分类器进行了研究，把 Iris 数据随机的平均分为互不相交的两组，分别设计分类器，通过研究这两个分类器的相似性，来判断新分类器是否稳定。

5.1 关于置信度的选取

随着置信度 α 由小到大，分类个数就会由少到多，即 α 越大类就越多， α 越小类就越少。那么 α 如何选取呢？针对文[44]中定义的置信度 α 的指标：

$$I_{\alpha} = \frac{\sum_{x \in \bigcup_{i=1}^n C_i} \mu_{\zeta_{bou}}(x)}{\sum_{x \in \bigcup_{i=1}^n C_i} \mu_{\zeta_{Total}}(x)}$$

我们注意到，指标并非对每个样本都作了考虑，所以修改为：

$$I'_{\alpha} = \frac{\sum_{x \in X} \mu_{\zeta_{bou}}(x)}{\sum_{x \in X} \mu_{\zeta_{Total}}(x)}$$

这里 $\zeta_{bou} = \bigvee_{i=1, j=1}^n (\zeta_{C_i} \wedge \zeta_{C_j})$ ， $\zeta_{Total} = \bigvee_{i=1}^n \zeta_{C_i}$ ， $i \geq 2$ 。

根据第四章提到的 AFS 模糊逻辑聚类分析算法及应用这种算法对 Iris 数据处理所得结果的基础上，通过 Iris 数据对置信度 α 的选取，我们做进一步的研究。

根据 M_{λ} ，选取不同的置信度 $\alpha \in [0, 1]$ ，会得到不同的聚类结果，在第四章提到的 AFS 模糊逻辑聚类分析算法的基础上，我们给出 α 与指标 I'_{α} 的曲线图和每个 α 所对应的分类结果的准确率（见图 5.1）。这里，分类结果准确率的算法如下定义：

定义 5.1 设在置信度 α 下将论域 X 分类为： C_1, C_2, \dots, C_l 类，正确的分类为： D_1, D_2, \dots, D_s 类，定义准确率为：

$$N = \sum_{i=1}^l \max\{m(C_i \cap D_1), m(C_i \cap D_2), \dots, m(C_i \cap D_s)\}$$

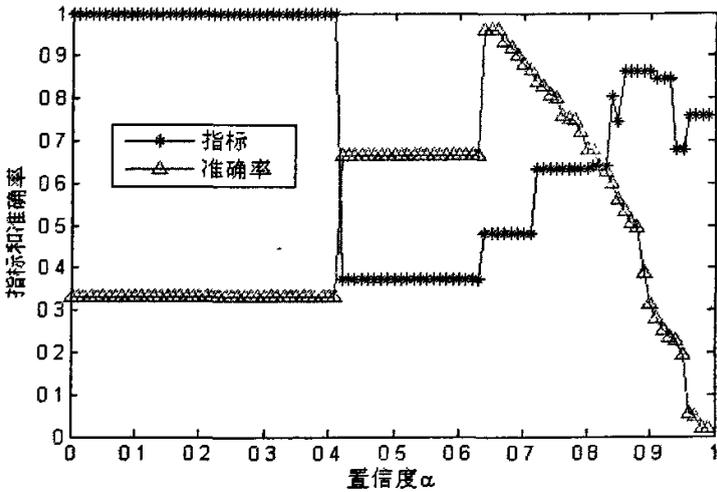


图 5.1 置信度 α 与指标 I'_α 的曲线图

Fig.5.1 curve figure of α and I'_α

注：当 $0 \leq \alpha \leq 0.42$ 时所有样本属于一类，当 $0.42 < \alpha \leq 0.6333$ 时所有样本分为两类：一类为 $\{x_1, x_2, \dots, x_{50}\}$ ，另一类为 $\{x_{51}, x_{52}, \dots, x_{150}\}$ ，当 $0.6333 < \alpha \leq 0.72$ 时，样本分为三类。其中 $0.6333 < \alpha \leq 0.65$ 时分类效果最好，只有五个样本没有分对，它们是 $x_{71}, x_{78}, x_{84}, x_{107}, x_{120}$ 。当 $\alpha > 0.85$ 时，因为很多样本隶属于它们自己的描述的程度低于 α ，所以这些样本就不再出现在分类结果中。

从图像上我们得出这样的结论：指标 I'_α 越小，分类结果越清晰， I'_α 越大，分类结果越混浊。当 I'_α 最小时，分类结果是最直观、最明显的。另外我们在图像中看到：在 0.64 处取得最佳分类结果，且 $\alpha = 0.64$ 时， $I'_\alpha = I'_{0.64} = 0.4797$ 在 I'_α 取得最小值 0.3750 的附近。由此可见，置信度 α 在 I'_α 的最小值附近选取时，分类结果最好。

5.2 基于 AFS 模糊逻辑的模糊分类器设计

第一步：根据所选训练样本建立 AFS 结构 (M, τ, X) 。

第二步：由第四章提到的 AFS 模糊逻辑聚类分析算法得出最佳分类和每一类的描述 ξ_{c_i} 。

第三步：对于测试样本 y ，计算它属于模糊概念 $\xi = \sum_{A \in EM} A$ 的隶属度：

$$\mu_{\xi}(y) = \sup_{A \in EM} (m(A(\{y\})) / m(X))$$

这里 $A(\{y\}) = \{x \in X \mid x \text{ 属于 } A \text{ 的程度小于等于 } y \text{ 属于 } A \text{ 的程度}\}$ 。

分别计算 $\mu_{c_i}(y)$ ，对于每个测试样本 y ， $\mu_{c_i}(y)$ 是 y 属于类 C_i 的程度， $i = 1, 2, \dots, c$ 。

下面利用给出的算法对 Iris 数据设计分类器，我们共做 10 次实验，随机选取 60%（其中每一类也是 60%）样本做训练样本，即随机选取第一类：30 个样本；第二类：30 个样本；第三类：30 个样本，共 90 个样本做为训练样本，余下的 40% 即 60 个样本做测试样本。

第一次试验：

第一步：根据所选的 90 个训练样本，建立 AFS 结构 (M, τ, X') 。这里 $M = \{m_1, m_2, \dots, m_8\}$ ，其中 m_1 为花瓣长度大即花瓣长， m_2 为花瓣长度不大即花瓣短， m_3 为花瓣长度适中， m_4 为花瓣长度不适中， m_5 为花瓣宽， m_6 为花瓣窄， m_7 为花瓣宽度适中， m_8 为花瓣宽度不适中。 X' 是从 X 中随机选取的 90 个训练样本（其中每一类选取 30 个）的集合，所选样本 x 的标号如下：

第一类：1, 2, 3, 4, 5, 7, 8, 9, 12, 13, 15, 20, 21, 23, 27, 28, 29, 30, 31, 33, 35, 36, 37, 38, 40, 41, 42, 46, 47, 48；

第二类：51, 52, 53, 58, 59, 61, 63, 64, 65, 66, 68, 69, 70, 71, 73, 74, 75, 77, 78, 79, 80, 82, 84, 85, 90, 91, 93, 94, 95, 100；

第三类：101, 102, 103, 104, 105, 106, 111, 114, 117, 118, 120, 122，

124, 125, 126, 128, 130, 131, 132, 133, 134, 136, 142, 143, 145, 146, 147, 148, 149, 150。

第二步：由第四章提到的 AFS 模糊逻辑聚类分析算法，对 X' 进行模糊聚类，得到三类描述：

$\xi_{c_1} = \{m_6\} + \{m_2, m_4\}$ ，即第一类为“花瓣窄的杉树”或“花瓣短且不适中的杉树”；
 $\xi_{c_2} = \{m_7\} + \{m_3\}$ ，第二类为“花瓣宽度适中的杉树”或“花瓣长度适中的杉树”；
 $\xi_{c_3} = \{m_1\} + \{m_5\}$ ，即第三类为“花瓣宽的杉树”或“花瓣长的杉树”。

第三步：分别计算 $\mu_{\xi_{c_1}}(y)$ ， $\mu_{\xi_{c_2}}(y)$ ， $\mu_{\xi_{c_3}}(y)$ 。下面给出 60 个测试样本在三类描述下的隶属度（见图 5.2）和 分类错误的样本（见表 5.1），实验中 60 个样本共分错 1 个。

第二次试验：

该实验得到的三类描述为： $\xi_{c_1} = \{m_6\} + \{m_2\}$ ， $\xi_{c_2} = \{m_7\} + \{m_3\}$ ， $\xi_{c_3} = \{m_1\} + \{m_5\}$ 。

下面给出 60 个测试样本在三类描述下的隶属度（见图 5.3）和分类错误的样本（见表 5.2），实验中 60 个样本共分错 2 个。

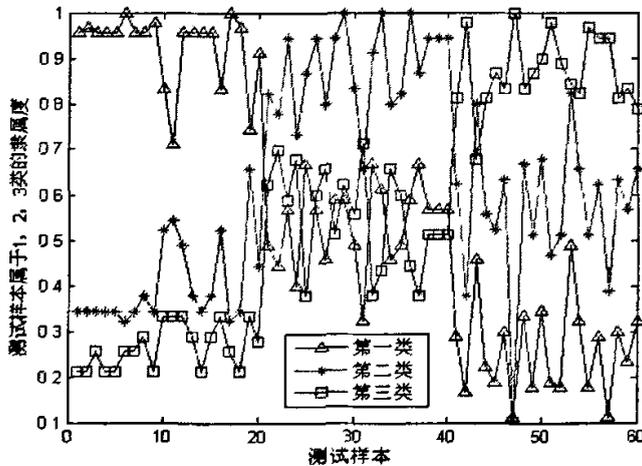


图 5.2 60 个测试样本属于 1, 2, 3 类的隶属度

Fig.5.2 Membership degree of 60 testing datas belong to 1, 2, 3 classes

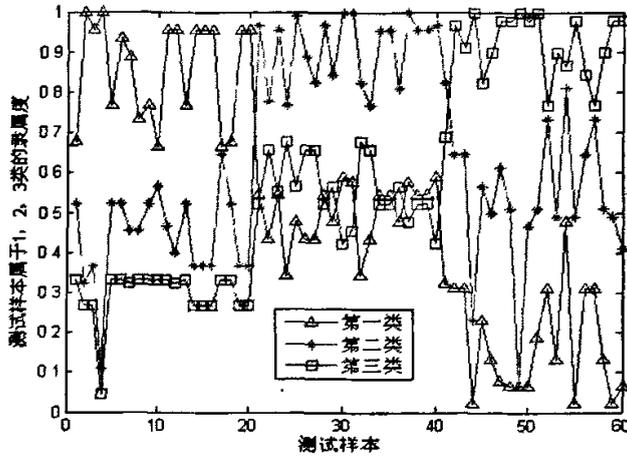


图 5.3 60 个测试样本属于 1, 2, 3 类的隶属度

Fig.5.3 Membership degree of 60 testing datas belong to 1, 2, 3 classes

表 5.1 分类错误的样本

Tab. 5.1 the samples of error clustered

测试样本标号	107
$\mu_{t_{c_1}}'(y)$	0.3222
$\mu_{t_{c_2}}'(y)$	0.8222
$\mu_{t_{c_3}}'(y)$	0.6889

表 5.2 分类错误的样本

Tab. 5.2 the samples of error clustered

测试样本标号	78	107
$\mu_{t_{c_1}}'(y)$	0.3222	0.4556
$\mu_{t_{c_2}}'(y)$	0.6556	0.8000
$\mu_{t_{c_3}}'(y)$	0.7111	0.6778

第三次试验：

该次试验得到三类描述为： $\xi_{C_1} = \{m_6\} + \{m_2\}$ ， $\xi_{C_2} = \{m_7\} + \{m_3\}$ ， $\xi_{C_3} = \{m_1\} + \{m_5\}$ 。

下面给出 60 个测试样本在三类描述下的隶属度（见图 5.4），60 个样本中没有分类错误的样本。其它试验这里就不一一列出了。下面给出十次试验中分类错误的个数（见表 5.3），这十次试验的平均准确率为 98%。对于测试错误的样本，由表 4.1 和表 4.2 我们可以看到，它属于那一类的隶属度并不明显大于它属于其它类的隶属度，这表明该测试样本确实是介于两类之间的。

表 5.3 十次试验中错误的样本

Tab. 5.3 the numbers of the samples of error clustered in ten tests

试验次数	1	2	3	4	5	6	7	8	9	10
错误个数	1	2	0	0	0	1	2	1	1	3

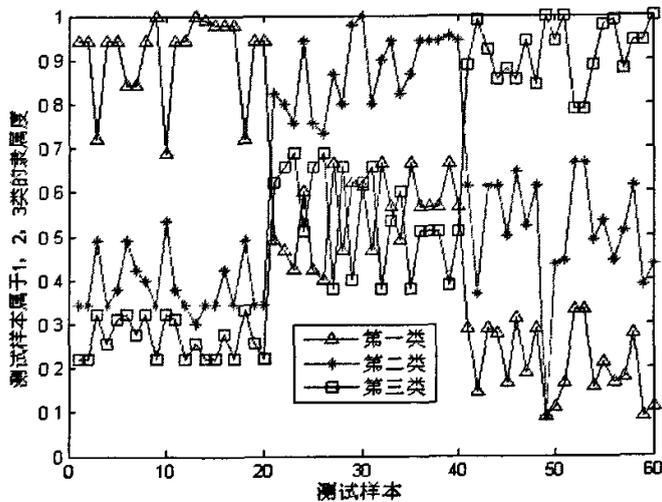


图 5.4 60 个测试样本属于 1, 2, 3 类的隶属度

Fig.5.4 Membership degree of 60 testing datas belong to 1, 2, 3 classes

注：1. 尽管在本例中， $X \subset \mathbb{R}^4$ ，但该算法可应用于样本的属性用各种数据类型或序关系描述的现实世界的问题。

2. 因为 M 中每个简单概念都有确切的语义，所以每类的模糊描述都表示确切的语义。

3. 通过这个实例可知 AFS 理论是研究人类智能、思维、推理、认识等重要的数学工具。

它为发现知识、表示知识研究现实世界中大规模复杂大系统提供了更加灵活且用力的框架。

5.3 分类器的稳定性分析

第一次试验：

我们随机的将 Iris 数据平均分为互不相交的两组，每组 75 个样本（其中每类选取 25 个），分别对其建立 AFS 结构，并由每组数据得到一个分类器，我们来研究这些分类器之间的相似性。随机选取 50% 的样本，建立一个 AFS 结构 (M, τ, X_1) ，余下的样本建立另一个 AFS 结构 (M, τ, X_2) ，分别在这两个 AFS 结构的基础上对 X_1 和 X_2 进行聚类分析，得到获取最佳分类结果时每一类的描述。

第一个结构 (M, τ, X_1) ，得到的描述为： $\zeta_{C_1} = \{m_6\} + \{m_2\}$ ， $\zeta_{C_2} = \{m_7\} + \{m_3\}$ ， $\zeta_{C_3} = \{m_1\} + \{m_5\}$ 。

第二个结构 (M, τ, X_2) ，得到的描述为： $\zeta_{C_1} = \{m_6\} + \{m_2\}$ ， $\zeta_{C_2} = \{m_7\} + \{m_3\}$ ， $\zeta_{C_3} = \{m_1\} + \{m_5\}$ 。我们发现两次试验对每类的描述是一致的。

我们接下来作三次和上面同样的试验，下面给出每次试验的描述。

第二次试验：

第一个结构 (M, τ, X_1) ，得到的描述为： $\zeta_{C_1} = \{m_6\} + \{m_2, m_4\}$ ， $\zeta_{C_2} = \{m_7\} + \{m_3\}$ ， $\zeta_{C_3} = \{m_1\} + \{m_5\}$ 。第二个结构 (M, τ, X_2) ，得到的描述为： $\zeta_{C_1} = \{m_6\} + \{m_2, m_4\}$ ， $\zeta_{C_2} = \{m_7\} + \{m_3\}$ ， $\zeta_{C_3} = \{m_1\} + \{m_5\}$ 。两次试验中每类的描述也是一致的。

第三次试验:

第一个结构 (M, τ, X_1) , 得到的描述为: $\xi_{C_1} = \{m_6\} + \{m_2\}$, $\xi_{C_2} = \{m_7\} + \{m_3\}$, $\xi_{C_3} = \{m_1\} + \{m_5\}$ 。第二个结构 (M, τ, X_2) , 得到的描述为: $\xi_{C_1} = \{m_6, m_8\} + \{m_2, m_4\} + \{m_2, m_6\}$, $\xi_{C_2} = \{m_7\} + \{m_3\}$, $\xi_{C_3} = \{m_1\} + \{m_5\}$ 。

第四次试验:

第一个结构 (M, τ, X_1) , 得到的描述为: $\xi_{C_1} = \{m_6\} + \{m_2\}$, $\xi_{C_2} = \{m_7\} + \{m_3\}$, $\xi_{C_3} = \{m_1\} + \{m_5, m_8\}$ 。第二个结构 (M, τ, X_2) , 得到的描述为: $\xi_{C_1} = \{m_6\} + \{m_2, m_4\}$, $\xi_{C_2} = \{m_7\} + \{m_3\}$, $\xi_{C_3} = \{m_1\} + \{m_5, m_8\}$ 。

我们看到, 这四次试验中对第一类的描述共有三种:

描述 1: $\{m_6\} + \{m_2\}$, 描述 2: $\{m_6\} + \{m_2, m_4\}$, 描述 3: $\{m_6, m_8\} + \{m_2, m_4\} + \{m_2, m_6\}$ 。

下面给出 150 个样本在这三种描述下的隶属度 (见图 5.5), 由图 5.5 可知每个样本属于这三种描述的程度都是很接近的, 这表明这三种描述在意义上是相似的。

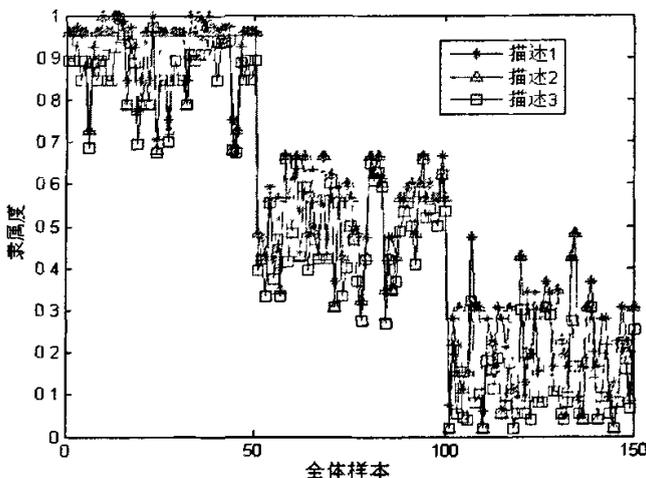


图 5.5 四次实验中关于第一类的三种描述的隶属函数

Fig.5.5 Membership function about the first class's three description in the four tests

四次试验中对第二类的描述都是一致的。四次试验中对第三类的描述共有两种：描述 1: $\{m_1\} + \{m_5\}$ ，描述 2: $\{m_1\} + \{m_5, m_8\}$ 。下面给出 150 个样本在这两种描述下的隶属度（见图 5.6），由图 4.11 可知每个样本在这两种描述下的隶属度也都是很接近的，这表明这两种描述在意义上是相似的。

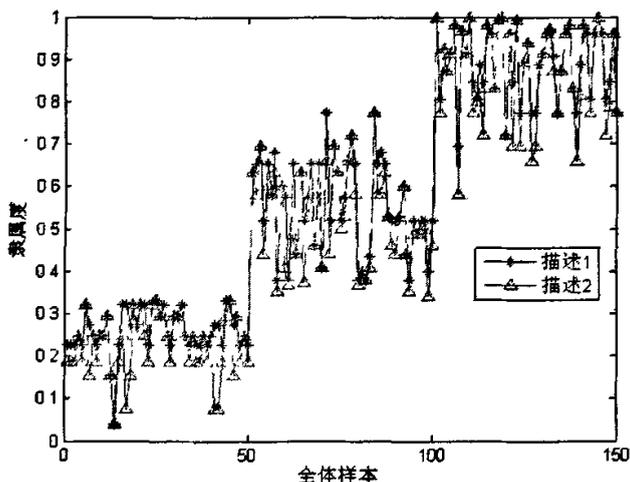


图 5.6 四次试验中关于第三类的两种描述的隶属函数

Fig.5.6 Membership function about the third class's two description in the four tests

通过这四次试验，发现所得到的八个分类器中对于同一类的描述是十分相似的，对于每一次试验中的两个分类器的描述也是相似的。也就是说，当我们选取的样本没有重复的时候，得到的分类器是相似的。这就说明我们的这种新的分类器是稳定的。

第6章 结论

6.1 本文研究的主要工作

通过对 AFS 模糊逻辑的聚类分析算法的研究, 我们不仅得到了新的分类分析算法, 而且很好的将其应用到 Iris 这种样本较多的数据集中。

本文在确定隶属度时, 只用到了样本属性上的序关系, 而没有用到样本属性上的数值。通过实际例子可以发现, 这种方法也能够得到很高的准确率。我们利用个改进后的 AFS 模糊逻辑聚类分析算法, 对 Iris 数据进行模糊聚类研究的结果, 表明了置信度 α 在 I_{α}' 的最小值附近选取时, 分类结果最好。

另外, 所设计的基于 AFS 模糊逻辑的模糊分类器不仅在本质上与以往的方法不同, 而且可以将其应用到那些样本属性只能用序关系描述而无法用数值描述的数据集。因为以往的分类器大多只是适用于样本属性用数值来描述的数据集, 所以对于样本属性用序关系描述的就非常困难或根本不可能处理。分类器设计过程的每一步都与人的认识过程类似, 由于分类器的表示形式具有明确的语义, 所以完全可被每个人所理解。

在我们设计的该分类器算法的第二步中, 关于每一类的模糊描述 ζ_C 可应用于大型模糊系统的描述和模式识别, 这种描述即可被人们理解又可应用许多深刻抽象的数学工具处理。事实上, 该分类器可被看做训练样本中蕴涵的知识表示, 即新的数据挖掘方法。我们对分类器稳定性的研究, 说明了这种新模糊分类器的算法是十分简洁和有效的。因为 AFS 结构与计算机的数据库在结构上很相似, 所以非常便于计算机操作。本文例子中的计算就是用我们编辑的 Matlab 程序计算的。基于 AFS 方法的模糊聚类分析是研究人类概念及其结构的重要数学方法, 并且应用它可使计算机为人类做更多的智能工作。

6.2 有待进一步研究的工作

AFS 理论是一个崭新的理论, 它能应用到很多领域中, 本文对 AFS 模糊逻辑聚类方法及其分类器算法的研究仅仅是初步的, 有待进一步研究和认识, 特别是它们结合其他方法 (如遗传算法、神经网络、决策树、案例算法等) 在数据挖掘

中的应用。对于它的研究越深入,对应用越有利,这个领域的研究前景是广阔的。作者深信,基于 AFS 模糊逻辑的神经网络、遗传算法值得期待。

参考文献

- [1] Zadeh L A, Fuzzy sets[J], Information And Control, 8(1965): 338-353
- [2] 杨伦标, 高英仪. 模糊数学原理及应用[M], 第二版, 1995, 华南理工大学出版社, 7-10
- [3] 陈貽源, 模糊数学[M]. 第一版, 1984年3月, 华中工业出版社, 13-20
- [4] 荃垆, 模糊数学[M]. 第一版, 1999年10月, 华中工业出版社, 13-20
- [5] 彭祖贇, 孙温玉, 模糊(fuzzy)数学极其应用[M], 第一版, 2002年3月, 武汉大学出版社, 2-10
- [6] 韩立岩, 汪培庄. 应用模糊数学[M], 第二版, 1998年1月, 首都经济贸易大学出版社, 2-42
- [7] 陈启浩. 模糊集及其在模糊推理中的应用[M], 第一版, 2000年8月, 北京师范大学出版社, 4-42
- [8] 王国俊. 广义拓扑分子格[J], 中国科学A辑, 1983, 12, 1063-1072
- [9] Wang Guo-jun. Theory of topological molecular lattices[J], Fuzzy Sets and Systems, 1992, 47, 351-376
- [10] Liu Xiaodong. The Fuzzy Theory Based on AFS Structure and AFS Algebra[M], Dalian Maritime University Press, 1998
- [11] Liu Xiaodong. The Fuzzy Sets and System Based on AFS Structure, EI Algebra and $E\pi$ algebra[J], Fuzzy Sets and Systems, 1998, 95, 179-188
- [12] Liu Xiaodong. A New Fuzzy Model of Pattern Recognition and Hitch Diagnosis of Complex Systems[J], Fuzzy Sets and Systems, 1999, 104, 289-296
- [13] Liu Xiaodong. The Fuzzy Theory Based on AFS Algebra and AFS Structure[J], Journal of Mathematical Analysis and Applications, 1998, 217, 459-478
- [14] Liu Xiaodong. The Topology on AFS Algebra and AFS Structure[J], Journal of Mathematical Analysis and Applications, 1998, 217, 479-489
- [15] Liu Xiaodong. The Structure of Fuzzy Matrices[J], Journal of Fuzzy Mathematics, 1994, 2, 311-325
- [16] Liu Xiaodong. Two Algebraic Structures of AFS Structure[J], Journal of Fuzzy

- Mathematics, 1995, 3, 561-562
- [17] 刘晓东, 张庆灵. AFS 模糊逻辑及其在聚类分析和模糊规则提取中的应用[J] 模糊系统与数学, 2002, 16 (1), 39-48
- [18] Liu Xiaodong. A New Mathematical Axiomatic System of Fuzzy Sets and Systems[J]. Journal of Fuzzy Mathematics, 1995, 3, 559-560
- [19] Kim K. H. Boolean matrix Theory and Applications[M], Inc.: Marcel Dekker, 1982
- [20] Andreja Tepavčević, Goran Trajkovski, L-fuzzy lattices an introduction[J], Fuzzy Sets and Systems, 2001, 123, 209-216
- [21] Ruspini E H, A new approach to clustering[J], Inf. cont., 1969, 15(1): 22-32
- [22] Bart Kosko, (黄崇明译), 模糊工程[M], 西安交通大学出版社, Prentice Hall, 1998
- [23] Jiawei Han, Micheline kamber, 范明, 孟小峰 译. 第一版, Data Mining: Concepts and Techniques[M], 2001, 机械工业出版社
- [24] AL-Sultan K S, Selim S Z, A simulated annealing algorithm for the clustering problem[J], Pattern Recognition, 1991, 24(10): 1003-1008
- [25] 赵丽萍, 数据库系统原理[M], 第一版, 2002年8月, 华东理工大学出版社, 1-20
- [26] 罗晓沛, 数据库技术[M], 第一版, 1999年9月, 清华大学出版社, 298-360
- [27] 刘同明等, 数据挖掘技术及其应用[M], 第一版, 2001年9月, 国防工业出版社, 1-20
- [28] Jae Dong yang, An image retrieval model based on fuzzy tripies[J], Fuzzy Sets and Systems 2001, 121, 459-470
- [29] 于加举, 王昕, 刘晓东. 概念的 EI 代数表示和布尔矩阵表示的研究[J], 大连海事大学学报, 2004, 30 (3): 95-99
- [30] J. L 凯莱 吴从忻等译. 一般拓扑学[M] 科学出版社 1982
- [31] 靳正大. 不分明拓扑空间引论[M] 大连海运学院出版社 1982
- [32] 熊金城. 点集拓扑学讲义[M] 高等教育出版社 1981
- [33] 蒲保明, 蒋继光, 胡淑礼. 拓扑学[M] 高等教育出版社 1985
- [34] 邱望仁, 丁蕊, 刘晓东, 张运杰. 拓扑分子格点的性质及其在概念分析中的应用[J] 大连海事大学学报 200403
- [35] 李凡. 一个采用模糊产生式规则的推理方法[J] 华中理工大学学报, 1992, 20(3):

145—150.

- [36] 李凡. 近似推理[M]. 北京:科学出版 1995. 131-1 71
- [37] 李凡. 一个有效的区间值模糊推理方法[J] 应用科学学报 vol. 18, No. 4 Dec.. 2000
- [38] 周晓越,王昕,刘晓东. 基于 AFS 方法的模糊推理规则[J] 大连海事大学学报 2003 29 (1) 78-83 2003 年 3 月
- [39] 周晓越, 刘晓东. 基于 AFS 理论和贴近度的模糊推理规则及属性约简 硕士学位论文 大连 大连海事大学 2004
- [40] 王昕, 刘晓东. The design of data query based on AFS fuzzy Algorithm proceedings of 2002 International conference on Management Science and Engineering ,Volume I 114-118 2002 年 10 月
- [41] 干毅兵, 刘晓东. AFS 结构分解算法及其在故障诊断中的应用 硕士学位论文 大连 大连海事大学 2004
- [42] X. Liu, W. Pedrycz and Q. Zhang, Axiomatic Fuzzy sets logic, IEEE International Conference on Fuzzy Systems, voll, pp 55-60, 2003.
- [43] X. Liu and W. Pedrycz, The Development of Fuzzy Decision Trees in the Framework of axiomatic Fuzzy Set Logic, Applied Soft Computing, accepted, 2005, available online.
- [44] X. Liu, W. Wang, T. Chai, The Fuzzy Clustering Analysis Based on AFS Theory, IEEE Transactions on Systems, Man and Cybernetics Part B, 2005, 35(5):1013-1027.
- [45] X. Liu, K. Zhu, H. Huang, The Representations of Fuzzy Concepts Based on the Fuzzy Matrix Theory and the AFS Theory, IEEE International Symposium on Intelligent Control, October 5-8, 2003, Texas, USA.
- [46] X. Liu, Q. Zhang, The Fuzzy Cognitive Maps Based on AFS Fuzzy Logic, Dynamics of Continuous, Discrete and Impulsive Systems, Series A: Mathematical Analysis, 2004, 11(5-6):787-796.
- [47] X. Liu, W. Liu, Credit Rating Analysis with AFS Fuzzy Logic, Lecture Notes in Computer Science, L. Wang, K. Chen, and Y.S. Ong (Eds.): 2005, LNCS 3612, pp. 1198-1204, Springer-Verlag Berlin Heidelberg. 22
- [48] X. Liu, T. Chai and W. Wang, Approaches to the Representations and Logic Operations for Fuzzy Concepts in the Framework of Axiomatic Fuzzy Set Theory I, II,

Information Sciences, in press, online available 2005.

[49] Y. J. Zhang, D. Q. Liang and S. C. Tong, On AFS algebra part I, Information Sciences, 2004, 167:263-286.

[50] Y. J. Zhang, D. Q. Liang and S. C. Tong, On AFS algebra part II, Information Sciences, 2004, 167:287-303.

[51] 任艳, 王显昌, 刘晓东, 基于AFS模糊逻辑的聚类分析方法研究(一) Proceedings of the 6th World Congress on Control and Automation, June 21 - 23, 2006, Dalian, China, 4244-4248.

[52] 王显昌, 任艳, 刘晓东, 基于AFS模糊逻辑的聚类分析方法研究(二), Proceedings of the 6th World Congress on Control and Automation, June 21 - 23, 2006, Dalian, China, 4199-4203.

攻读学位期间公开发表的论文

[1] 王显昌,任艳,刘晓东, 基于 AFS 模糊逻辑的聚类分析方法研究(二), Proceedings of the 6th World Congress on Control and Automation, June 21 - 23, 2006, Dalian, China, 4199-4203, EI 检索。

[2] 王显昌,丁瑞,刘晓东, New Approaches to the Fuzzy Classification via AFS Theory, International Journal of Information & Systems Sciences . 已接收。

[3] 王显昌, 刘晓东, 任艳, The Improved Fuzzy Clustering Algorithm Based on AFS Theory and Its Applications to Wisconsin Breast Cancer Data, 投于 icic07, SCI 检索。

[4] 刘晓东, 王显昌, A Novel Approach to Fuzzy Clustering Based on AFS Theory, 投于 Fuzzy Sets and System, SCI 检索。

[5] 任艳, 王显昌, 刘晓东, 基于 AFS 模糊逻辑的聚类分析方法研究(一) Proceedings of the 6th World Congress on Control and Automation, June 21 - 23, 2006, Dalian, China, 4244-4248, EI 检索。

[6] 任艳, 刘晓东, 王显昌。 The Improved Fuzzy Clustering Algorithm Based on AFS Theory and Its Applications to Wisconsin Breast Cancer Data, 投于 icic07, SCI 检索。

致 谢

回顾近三年的研究生学习历程,我要由衷感谢我的导师刘晓东教授对我的关怀与悉心指导。刘老师严谨的治学态度,广阔的学术视野,思考问题时独特的视角,解决问题时对和谐与完美的追求,还有对我论文选题、写作方面的指导给了我许多启迪,令我终身难忘。他对复杂问题所独具的敏感直觉,使我避免了许多弯路;他的勤奋、执着为我树立了作为一名学者的典范,值得我尽毕生的努力去学习。刘老师不仅在学术上为我付出了大量的心血,而且在工作生活中也给了我许多关心和帮助。他的帮助犹如我人生道路上的一盏明灯,指引我顺利地走过了三年的研究生旅途,也将指导我今后的工作。

同时也要感谢张运杰教授、赵连昌教授、刘巍教授、张强副教授、杨晓光副教授等老师在这三年中给予我学习及生活中的无私帮助与关怀。最后,我要感谢师姐妹、师兄弟们及同学们给予我的关心和热情帮助。

研究生履历

姓 名	王显昌
性 别	男
出生日期	1980年10月22日
获学士学位专业及门类	数学 理学学士
获学士学位单位	辽宁师范大学
获硕士学位专业及门类	应用数学 理学硕士
获硕士学位单位	大连海事大学
通信地址	辽宁省大连市凌海路1号
邮政编码	116026
电子邮箱	wxcixll@sohu.com