

# 粒度聚类算法研究

徐丽<sup>1</sup> 丁世飞<sup>1,2</sup>

(中国矿业大学计算机科学与技术学院 徐州 221116)<sup>1</sup>

(中国科学院计算技术研究所智能信息处理重点实验室 北京 100080)<sup>2</sup>

**摘要** 信息粒度是对信息和知识细化的不同层次的度量。基于信息粒度的聚类分析方法,凭借能够灵活选择粒度结构,消除聚类结果和先验知识之间的不协调性,有效完成聚类任务等优点,成为国内外学者的研究热点之一。从粗糙集、模糊集、商空间3个理论角度与传统聚类算法相结合,阐述并分析了把粒度的思想引入到聚类中的有效算法及其优缺点,并对这样结合后处理高维复杂数据的可行性及有效性做了分析与展望。

**关键词** 信息粒度,粗糙集,模糊集,商空间理论,聚类算法

中图分类号 TP18 文献标识码 A

## Research on Granularity Clustering Algorithms

XU Li<sup>1</sup> DING Shi-fei<sup>1,2</sup>

(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China)<sup>1</sup>

(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China)<sup>2</sup>

**Abstract** Information granularity is a measure of different levels for refining information and knowledge. With the advantages of selecting granularity structure flexibly, eliminating incompatibility between clustering results and priori knowledge, completing clustering task effectively, granularity clustering methods become one of the focus at home and abroad. In this paper, combined the traditional clustering algorithms from the view of rough set, fuzzy set and quotient space theories, effective clustering algorithms with the idea of granularity and their merits and faults were studied and generalized. Finally, the feasibility and effectiveness of handling high-dimensional complex massive data with combination of these theories were forecasted and outlooked.

**Keywords** Information granularity, Rough set, Fuzzy set, Theory of quotient space, Clustering algorithm

## 1 引言

聚类是一种常见的数据分析工具,它就是把一些对象分成若干类,使得类间对象差异尽可能地大而类内对象差异尽可能地小,发现样本点间最本质的“抱团”性质<sup>[1]</sup>。聚类是数据挖掘、模式识别等众多领域的重要研究内容之一,在识别数据的内在结构方面具有极其重要的作用。作为一种无监督的机器学习方法,聚类分析在自然科学和社会科学中得到广泛应用。在信息检索和数据挖掘的过程中,聚类处理对于建立高效的索引、快速而准确的定位信息和进一步挖掘隐藏的知识具有重要的理论和现实意义。在图像分割和机器视觉、卫星图片分析、金融数据分类、医学图像的自动检测、生物考古等领域,聚类分析也起到了不可或缺的作用<sup>[2]</sup>。

随着社会科技的发展,传统的聚类方法已经不能满足分析求解当今数据的要求,海量的数据产生了“数据爆炸”,高维属性使其处理起来更加困难。探索新的聚类算法成为近年来国内外迫在眉睫的研究热点之一。学者们发现,粒度和聚类具有天然的相通性,引进粒度分析理论能够有效地完成聚类

任务。聚类分析是以“最优”相似度函数为基础,在所有可能的粒度中,求出一个“最优”粒度。从信息粒度的角度看,聚类就是在一个统一的粒度下分析和处理问题。把粒度计算的思想运用到聚类算法中,会得到较理想的结果。

人们分析问题时往往从不同角度、不同层次出发,遵循由表及里、层层深入的研究方法。凭借经验或专业知识把复杂的信息按照各自的特征和性能分成不同的块,通过对简单块的分解与合并,进而综合全面地认识事物并解决问题。这和粒度的本质思想<sup>[3,4]</sup>是一致的。粒度思想模仿人类思考问题的方式,把问题粒化后,经过粗细粒度的反复迭代,用低成本的、足够满意的近似解替代精确解,从而更好地认识和刻画世界。

在聚类算法中,相似度函数是决定聚类结果的主要因素之一<sup>[5]</sup>。通过改变它,使得聚类的处理对象变换到不同的粒度空间。由于粒度的变换可以将复杂的问题简单化,即可将对象简化成若干个保留重要特征和性能的点以便于分析,因此可以通过粒度的粗细变化来进行聚类。聚类的实质<sup>[6,7]</sup>就是在样本点之间定义一种等价关系,一个等价关系定义了一

到稿日期:2010-09-01 返修日期:2011-01-14 本文受国家自然科学基金(60975039),江苏省基础研究计划(自然科学基金)(BK2009093)资助。

徐丽(1986-),女,硕士生,主要研究方向为聚类分析、粒度计算, E-mail: xl412@126.com; 丁世飞(1963-),男,博士,教授,博士生导师,主要研究方向为人工智能、数据挖掘、粒度计算。

个划分,这个划分把样本点化成若干子集,一个子集就对应着聚类形成的一个簇。每个簇中任意两个样本点性质相近,在当前给定的阈值下没有区别。根据具体问题得到不同的等价关系,每个等价关系都对应着一个划分,从粒度的角度看,就是变换到了不同的粒度空间。在粗细粒度空间“跳来跳去”,就得到了最终需要的聚类结果。

在实际应用中,常常存在聚类结果和先验知识不协调的缺陷,粒度聚类算法可通过变换不同粒度下的特征空间来消除这种不协调性。此外,传统的聚类算法还存在对孤立点敏感、对参数设置敏感等不足,引入粒度的思想并采用两者相结合的方法,可弥补这种不足,从而提高算法的效率。近年来,有关粒度聚类算法的研究论文层出不穷,主要应用于对传统聚类算法的改进<sup>[8-10]</sup>、图像处理<sup>[11]</sup>、生物学中的进化计算<sup>[12]</sup>、Web 页面文本聚类<sup>[13-15]</sup>、神经网络<sup>[16]</sup>等众多领域。

## 2 粒度聚类算法

### 2.1 基于粗糙集的聚类算法

粗糙集理论<sup>[17,18]</sup>作为一种能够定量分析处理不精确、不一致、不完整等各种不完备信息的有效数学工具,是粒度计算的理论模型之一。凭借其数学基础成熟、不需要先验知识以及易用性等优点,将其应用于聚类分析能够提高聚类算法的效率和聚类结果的准确性。

不可分辨关系是粗糙集理论中的一个关键概念,它通常和一个属性集合相联系。给定一个知识表达系统  $S=(U, R, V, F)$ ,对于每个属性子集  $B \subseteq R$ ,定义如下不可分辨关系。

$$IND(B) = \{(x, y) | (x, y) \in U^2, \forall b \in B(b(x) = b(y))\}$$

针对不可定义集,无法构造一个公式来精确描述,只能通过上下界逼近的方式来刻画,这就是粗糙集理论中的上下近似集。

对于每个子集  $X \subseteq U$  和不可分辨关系  $B$ ,  $X$  的上近似集  $B^*(X)$  和下近似集  $B_*(X)$  定义为

$$B^*(X) = \cup \{Y_i | (Y_i \in U | IND(B) \wedge Y_i \subseteq X)\}$$

$$B_*(X) = \cup \{Y_i | (Y_i \in U | IND(B) \wedge Y_i \cap X \neq \emptyset)\}$$

当且仅当  $B^*(X) \neq B_*(X)$  时,称集合  $X$  为关于关系  $B$  的粗糙集。

大部分的聚类算法都是基于构造对象之间的相异度矩阵,但实际上数据都是用数据矩阵的形式表现的,所以需要从数据矩阵向相异度矩阵转换,这必然会产生噪声和失真。采用粗糙集的方法(RSDC)<sup>[19]</sup>,直接用数据矩阵进行聚类,根据相关理论知识对数据不必要的属性进行约简,从而体现了数据固有的差别。此外,该算法对符号属性和数值属性也都有良好的聚类结果。但不足的是它只适用于对象个数和属性个数都不多的小样本聚类。

在处理图像分割时,K-均值算法是常用的聚类方法。它有一定的自适应性,能够实现动态聚类,但结果易受初始聚类中心的影响。同时,由于图像信息的复杂性和相关性,在处理过程中常出现不完整性和不精确性问题。因此结合粗糙集理论,先用 K-均值算法将原图像划分为初始的几个类后,再引入粗糙集中的不可分辨关系,将图像划分为若干小区域,然后采用小区域中不同点约简后的属性分配权值来计算类间的相异度。最后,由不同点决定的等价关系得到满意的分割结果<sup>[11]</sup>。

在信息论中,熵表示的是不确定性的度量。利用粗糙相似关系,基于熵的概念提出一种聚类算法(EIGRSCA)<sup>[20]</sup>,其

通过计算每个数据点的熵来选取聚类中心。根据越是有有序排列的数据,熵值越小的信息论原理,在聚类过程中,不需要调整所有数据点的熵值,这样大大降低了计算的复杂度,提高了算法的有效性。粗糙集最大的特点是不需要任何先验知识来处理数据,用它进行数据预处理,可以精简数据的结构。鄂旭等人<sup>[21]</sup>提出了一种聚类算法,首先用粗糙集进行属性约简,然后用超立方的概念对属性进行泛化。实验证明,此算法不受数据空间分布特征的影响,可以根据特征值排除孤立点且只扫描一遍信息表即可得到聚类划分,执行效率较高。由于数据的每个属性特征在聚类分析过程中的重要性是不一样的,因此提出一种基于粗糙集的加权聚类算法<sup>[22]</sup>,它首先用粗糙集相似模型理论求出初始等价类,然后对初始等价类进行削减,将原始系统扩展为决策系统后再利用信息熵来评价属性的重要性。以属性的重要度为权值对每个属性进行加权,重新考虑对象的相似关系,重新聚类。通过加权,该算法在处理大数据集的复杂情况下,优于传统算法。由此看来,属性约简是粗糙集研究的一个很实用、很重要的方向。

Leader 算法<sup>[23,24]</sup>是运用广泛的层次聚类算法,它用一个单阈值来划分样本,这样的硬划分没有考虑到聚类分析中固有的不确定性。苗夺谦提出一种加入粗糙集和粒计算的思想<sup>[9]</sup>,其借鉴粗糙集中上下近似的概念,在经典的 Leader 算法基础上,增加一个阈值,使用双阈值来替代簇的上下边界,把处于边界区域(类别间产生重叠的部分)中的数据先缓存起来,等到其他的数据集包含了较充分的划分信息后再对其进行迭代聚类。然后再用粒度不断细化的思想,构造出一些包含很少样本的簇,以利于发现孤立点。

谱聚类<sup>[25]</sup>是根据图的谱分割原理提出的一种聚类算法,该算法与数据的维数无关,执行容易,具有识别非凸分布聚类的能力,因此在实际问题中有着广泛的应用。但现有的谱聚类结果多为精确集,而对真实数据集中存在的重叠现象却无法识别。苗夺谦等人<sup>[26]</sup>把粗糙集的思想应用到谱聚类中进行扩展,使得聚类结果成为具有上下近似定义的、类之间存在重叠区域的结构,从而真实地反映了现实数据固有的模糊性质。另外,对孤立点的有效处理减小了类内距离,从而体现了数据对象的“抱团”性质。K-modes<sup>[27]</sup>是对经典 K-means 算法的改进,不仅能处理数值数据还能处理符号数据。它用简单匹配方法度量对象之间的距离,没有考虑到同一个属性下两个不同值之间的相似性。梁吉业等人<sup>[8]</sup>应用粗糙集理论中的上下近似,从属性本身和其他相关属性两个角度出发,提出一种新的相似性度量。考虑到字符属性,还改进了类中心的表示方法,以充分反映类的特征,使其能够体现属性在该类上的分布,有效地提高聚类精度。

### 2.2 基于模糊集的聚类算法

Zadeh<sup>[28]</sup>提出的模糊集理论为现实中数据的软划分提供了强有力的分析工具,其中隶属函数是表达事物模糊性的重要概念。人们开始用模糊的方法来处理聚类问题,模糊聚类给出了样本属于各个类别的不确定程度的描述。由于模糊聚类能够有效地对类与类之间有交叉的数据集进行聚类,更贴近于现实世界,因此许多学者对模糊聚类算法进行了深入的研究。

蚂蚁算法<sup>[29]</sup>是研究聚类分析的有效算法,它模仿蚂蚁的觅食行为来对数据进行聚类。蒋志为等人<sup>[30,31]</sup>发现蚁群算法的基本模型和 Lumer Faieta 算法存在一些缺陷,会导致不相似的数据对象本该被拾起而未被拾起,相似的对象本该被

放下而未被放下,从而影响了聚类效果。其实“相似”本身就是一个模糊的概念,每个人对模糊事物的界限是不完全一样的,因此提出用模糊集理论知识来解决,基于平均距离定义了“相似”这一模糊子集的隶属函数,最终数据对象的拾起与放下由隶属度与置信水平 $\lambda$ 相比较来决定。通过调整 $\lambda$ 来降低对相似性参数 $\alpha$ 的敏感度,此算法还减少了LF算法中的参数数量,并降低了计算的复杂度,是对蚁群聚类算法的一个很好改进。

模糊C均值聚类算法(FCM)是经典的聚类算法,在很多领域得到很好的运用。但该算法默认所有的特征向量在聚类过程中作用相同,这就会干扰数据的分类,影响结果的准确度。基于此,提出了一些改进的方法,如提取相对重要的特征来聚类,但这需要一定的实际经验;使用优化特征权重来动态聚类,却只能发现团状的数据分类。因此,综合上述存在的不足,提出一种动态加权模糊核聚类算法<sup>[32]</sup>,即在对数据没有任何先验知识的情况下,利用核聚类的核心思想<sup>[33]</sup>来非线性映射 Mercer 核函数,把特征向量映射到高维空间,从而能够非线性划分非团状数据。同时根据 FCM 理论和动态调整权值,消除噪声特征向量对聚类的影响,使得目标函数达到最小,得到较好的聚类结果。但该算法引入了核函数,使得算法的聚类时间有所增加。

针对传统K-均值算法对初始聚类中心敏感的缺点,不仅结合粗糙集理论可以改进,利用模糊粒度的思想也能得到较好的聚类结果。张霞等人<sup>[34]</sup>提出一种优化初始聚类中心的新方法:根据模糊粒度理论,定义一个归一化的距离函数,用此函数对所有距离小于粒度 $d_i$ 的数据对象进行聚类,从而计算得到一组优化的初始聚类中心。该算法消除了K-均值算法对初始输入的敏感性,提高了算法的稳定性和准确性。

在传统的网格密度聚类算法上,用模糊集的隶属度对要聚类的基本区域及模糊扩展区域内的数据点进行计数<sup>[35]</sup>,考虑相邻网格对当前网格内数据点的影响,先用了属性相关分析的方法对不必要的属性进行约简,然后通过网格划分和计数来聚类。此算法有效缓解了不平滑聚类以及聚类边界模糊的问题,良好的属性关系可以通过粗糙集的属性约简来得到。邱保志等人<sup>[36]</sup>又提出一种基于网格和密度权值的模糊C均值算法,它也是对初始聚类中心的优化与改进,使用网格代表点也就是网格单元中所有数据点的平均值来代替原算法中的网格凝聚点,更好地体现了数据的分布,同时利用密度加权消除了噪声样本对聚类结果的影响。

### 2.3 基于商空间的聚类算法

商空间理论<sup>[37]</sup>是我国张铃教授和张钹院士首先提出的,它将不同的粒度世界与数学上的商集概念统一起来。用一个三元组 $(X, f, T)$ ,即论域、属性、结构来描述一个问题。在其论域上引入等价关系 $R$ ,通过等价关系 $R$ 构造等价类,得到商集 $[X]$ ,然后将 $[X]$ 当作新的论域,原问题就相应地变为 $([X], [f], [T])$ ,再在新的粒度空间进行问题的分析与求解。通过不同的等价关系构造不同的粒度世界,从而可以把问题简化,以便分析处理。在聚类分析中引入商空间的概念,就是在所有可能粒度中求出一个最优粒度。

在具体的聚类过程中,有的时候会遇到簇内的粒度相同且簇间的粒度也存在相似性的情况。为了遵循“类内相似,类间相异”的聚类准则,提出了一种基于商空间粒度的覆盖聚类算法<sup>[38]</sup>,它首先用交叉覆盖算法<sup>[39]</sup>得出聚类结果,通过比较类间的相似度来进行合并,最后采用粒度粗细变化来调整得

到的聚类结果。实验证明此算法适合高维数据聚类,更易于得到全局最优解,降低了计算复杂度,提高了聚类速度。王伦文等人<sup>[16]</sup>采用构造神经网络和商空间粒度相结合的聚类方法,既发挥了构造神经网络计算复杂度低的优点,也利用了商空间理论选取最优粒度聚类,对大规模复杂数据聚类具有时效性。基于商空间理论和聚类结合的算法很多,大都是用较成熟的算法得到聚类结果,然后再融入等价粒度的思想反复调整得到最优解。传统的模糊聚类方法计算相似矩阵时,需要大量的存储空间,徐峰等人<sup>[40-41]</sup>用商空间理论把论域进行合成,提出了基于 Gaussian 函数的模糊聚类算法,即采用距离度量表示信息粒度,并以信息融合技术从不同角度和层次来合成聚类结果,改变了传统意义上聚类是在一个统一的均匀粒度上描述样本集,通过对商空间理论中论域进行合成,确定最优粒度,降低了计算复杂度和时间复杂度。该算法不需要相似度矩阵,节省了空间,适合大样本聚类。唐旭清等人<sup>[42]</sup>也基于商空间理论,提出一种基于模糊相似关系和归一化距离的聚类方法,用以解决复杂系统的数据分析问题。

### 2.4 基于3大理论间融合的聚类算法

近年来,粒度计算3大理论的融合成为国内外学者的研究热点。虽然三者对于问题思考的出发点和解决的具体方法不一样,但它们有一个很大的共同点,就是都模仿了人类智能,从不同层面、不同角度采用粒度的思想去解决问题。如果把对它们取长补短、交汇融合后得到的理论方法运用到聚类分析中去,将会在效率上得到很大的改进和突破。

针对粗糙聚类算法对数据比例变换鲁棒性较差的问题,在粗糙聚类的框架下融合模糊聚类的思想,提出了一种带模糊权的粗糙聚类算法<sup>[43]</sup>,它把临界区域中对象的模糊隶属度作为调整聚类中心的作用权值,不仅改善了数据比例变换的鲁棒性,而且克服了传统粗糙聚类算法对阈值过于敏感的缺点。

多学科、多理论交叉融合是现今处理问题时常用的解决方法。随着数据变得越来越复杂,为了挖掘隐藏的有用信息,通常采用和多种方法结合的聚类算法,再融入信息粒度的理论思想,将会得到更加准确有效的信息知识。郝晓丽等人<sup>[12]</sup>提出了一种基于动态粒度的并行人工免疫聚类算法,即利用算法中的趋同、异化和传优算子,与动态粒度模型相结合。在粒度变化过程中,通过对粒度粗细粒化的调整,从并行优化和信息粒度的角度构造一种新的聚类方法来解决。支持向量机是一种基于统计学理论的机器学习方法,为了提高在大规模数据集上的训练效率,提出一种基于粒度聚类的模糊支持向量机算法<sup>[44]</sup>,它不仅降低了计算的时间复杂度,而且在提高训练效率的同时也保持了较好的分类结果。随着粒度思想研究的深入,基于粒度计算的聚类分析方法会具有更好的可行性和实用性。

在科学技术飞速发展的今天,高维数据在日常生活中占据着越来越重要的比例。受“维度效应”的影响,许多在低维数据空间表现良好的聚类方法运用在高维空间上往往无法获得好的聚类效果。近年来,为了解决高维数据聚类,较有效的方法是先降维再聚类<sup>[45-48]</sup>。选择合适的子空间,在低维度下进行子空间聚类。典型的方法有特征转换法和特征选择法。由于这也是基于距离的聚类,因此为了提高计算效率,需要抽取样本,但这样做可能会忽略掉对聚类结果有影响的对象元素,效果不是很理想。高维数据的稀疏性、空空间现象以及维灾现象使得针对高维数据的聚类算法面对更多的困难。因此高维数据聚类成为聚类分析的难点和重点。

**结束语** 粒度聚类算法在处理不精确、不完备、模糊的数据时具有优越性。本文在阐述聚类中的粒度原理的基础上,分析并概括了近年来把粒度思想运用于聚类分析的有效算法,并对近年来基于粗糙集、模糊集和商空间理论以及交叉融合的聚类算法做出了阐述和优缺点评价。

对于高维空间聚类,在做降维处理时,粗糙集是不错的属性约简方法,最后可采用数据融合的方法得到最终聚类结果。由于信息粒度的多样性,会导致对问题的多种不同的表示和解释的产生,增加了理解的难度,因此寻找高效的问题求解算法,最终实现构建统一的平台、基于信息粒度的理论模型交叉融合的有效聚类算法,是需要进一步研究和待解决的问题。

### 参 考 文 献

[1] 孙吉贵,刘杰,赵连宇. 聚类算法研究[J]. 软件学报,2008,19(1):48-60

[2] 李明华,刘全,刘忠. 数据挖掘中聚类算法的新发展[J]. 计算机应用研究,2008,25(1):13-17

[3] 李道国,苗夺谦,张红云. 粒度计算的理论、模型与方法[J]. 复旦学报:自然科学版,2004,43(5):837-841

[4] 苗夺谦,王国胤,刘清,等. 粒计算:过去、现在与展望[M]. 北京:科学出版社,2007

[5] 王伦文. 聚类的粒度分析[J]. 计算机工程与应用,2006,42(5):29-31

[6] 东波,白硕,李国杰. 聚类/分类中的粒度原理[J]. 计算机学报,2002,25(8):810-815

[7] 陈洁,张燕平,张铃,等. 基于信息粒度的聚类分析及其应用[J]. 中国图像图形学报,2007,12(1):87-91

[8] 白亮,梁吉业,曹付元. 基于粗糙集的改进 K-MODES 聚类算法[J]. 计算机科学,2009,36(1):162-176

[9] 殷钢,苗夺谦,段其国. 一种新的粗糙 Leader 聚类算法[J]. 计算机学报,2009,36(5):203-205

[10] 刘少辉,史忠植. 一种基于 Rough 集的层次聚类算法[J]. 计算机研究与发展,2004,41(4):552-557

[11] 刘岩,岳应娟,李言俊. 基于粗糙集的图像聚类分割算法研究[J]. 红外与激光工程,2004,33(3):300-302

[12] 郝晓丽,谢克朋. 基于动态粒度的并行人工免疫聚类算法[J]. 计算机工程,2007,33(23):194-196

[13] 钟茂生. Web 页面的模糊聚类[J]. 华东交通大学学报,2004,21(5):59-62

[14] Zheng Shang-zhi, Zhao Xiao-long, Zhang Bu-qun. Web Document Clustering Research Based on Granular Computing[C]// 2009 Second International Symposium on Electronic Commerce and Security. 2009:446-450

[15] Zhang Xia, Yin Yi-xin, Xu Ming-zhu. Research of Text Clustering Based on Fuzzy Granular Computing[C]// 2009 Second IEEE International Conference on Computer Science and Information Technology. 2009:288-291

[16] 徐银,周文江,王伦文. 基于构造型神经网络和商空间粒度的聚类方法[J]. 计算机工程与应用,2007,43(29):165-167

[17] Pawlak Z. Rough sets[J]. International Journal of Information and Computer Science,1982,11(5):341-356

[18] 王国胤,姚一豫,于洪. 粗糙集理论与应用研究综述[J]. 计算机学报,2009,32(7):1229-1246

[19] 孙士保,赵文涛,秦克云,等. 基于粗糙集的数据聚类方法研究[J]. 计算机工程与应用,2006,42(22):140-142

[20] 何明,冯博琴. 基于熵和信息粒度的粗糙集聚类算法[J]. 西安交通大学学报,2005,39(4):343-346

[21] 鄂旭,高学东,陈益,等. 一种基于粗糙集的聚类算法[J]. 计算机工程,2007,33(4):14-16

[22] 田慧,刘希玉,李章泉. 一种基于粗糙集的加权聚类算法[J]. 微计算机信息(管控一体化),2008,9(3):239-241

[23] Spath H. Cluster Analysis Algorithms for Data Reduction and Classification of Objects[M]. Chichester: Ellis Horwood Limited,1980

[24] Asharaf S, Murty M N, Shevade S K. Rough Set-based Incremental Clustering of Interval Data[J]. Pattern Recognition Letters,2006,27(6):515-519

[25] Bach R, Jordan M I. Learning spectral clustering [R]. University of California at Berkeley,2003

[26] 郑吉,苗夺谦,等. 一种基于粗糙集理论的谱聚类算法[J]. 计算机学报,2009,36(5):193-196

[27] Huang Zhe-xue. Extensions to the k-Means algorithm for clustering large data sets with categorical values[J]. Data Mining and Knowledge Discovery,1998,1(2):283-304

[28] Zadeh L A. Fuzzy logic=Computing with Words [J]. IEEE Trans. on Fuzzy Systems,1996,1(2):103-111

[29] Dorigo M, Maniezzo V, Colomni A. Ant System: Optimization by a Colony of Cooperating Agents[J]. IEEE Trans. on Systems, Man, and Cybernetics-Part B,1996,26(1):29-41

[30] 蒋志为. 基于模糊集的蚁群聚类算法研究[D]. 成都:西南交通大学,2006

[31] 朱树人,匡芳君,等. 基于粒度原理的蚁群聚类算法[J]. 计算机工程,2005,31(23):162-163

[32] 李颖,李传龙,马龙,等. 动态加权模糊核聚类算法[J]. 计算机工程与设计,2009,30(24):5584-5587

[33] 张莉,周伟达,焦李成. 核聚类算法[J]. 计算机学报,2002,25(6):587-590

[34] 张霞,王素贞,尹怡欣,等. 基于模糊粒度计算的 K-means 聚类算法研究[J]. 计算机科学,2010,37(2):209-211

[35] 杨悦,张健沛,等. 一种高维空间数据的模糊聚类算法[J]. 哈尔滨工程大学学报,2006,27(Suppl):485-188

[36] 邱保志,卢海艇. 基于网格和密度权值的模糊 C 均值聚类算法[J]. 计算机工程与设计,2010,31(4):822-824

[37] 张钹,张铃. 问题的求解理论及应用[M]. 北京:清华大学出版社,2007

[38] 闵鹏鹏,郑诚,胡哲. 基于商空间粒度的覆盖聚类算法的研究[J]. 计算机与现代化,2008,1(11):62-65

[39] 赵姝,张燕平,张铃. 基于粒度计算的覆盖算法[J]. 计算机科学,2008,35(3):225-227

[40] 徐峰,张铃,王伦文. 基于商空间理论的模糊粒度计算方法[J]. 模式识别与人工智能,2004,17(4):424-429

[41] 徐峰,张铃. 基于商空间的非均匀粒度聚类分析[J]. 计算机工程,2005,31(3):26-28

[42] 唐旭清,朱平,等. 基于模糊商空间的聚类分析方法[J]. 软件学报,2008,19(4):861-868

[43] 李订芳,章文,何炎祥. 一种新的带模糊权的粗糙聚类算法[J]. 信息与控制,2006,35(1):120-125

[44] 祁立,刘玉树. 基于聚类粒度的模糊支持向量机[J]. 哈尔滨工程大学学报,2006,27(Suppl)

[45] 陈慧萍,王煜,王建东. 高维数据挖掘算法的研究与进展[J]. 计算机工程与应用,2006,42(24):170-173

[46] 单世民,王新艳,张宪超. 高维分类属性的子空间聚类算法[J]. 小型微型计算机系统,2009,30(10):2016-2020

[47] 陈建斌. 高维聚类知识发现关键技术及应用[M]. 北京:电子工业出版社,2009

[48] 贺玲,蔡益朝,杨征. 高维数据聚类方法综述[J]. 计算机应用研究,2010,27(1):23-31