基于动态邻域的三支聚类分析

王平心[1,3] 刘 强[2] 杨习贝[2] 米据生[3]
(江苏科技大学理学院 江苏 镇江 212003)[1] (江苏科技大学计算机科学学院 江苏 镇江 212003)[2]
(河北师范大学数学与信息科学学院 石家庄 050024)[3]

摘 要 目前,大多数聚类方法是二支聚类,即对象要么属于一个类,要么不属于一个类,聚类的结果必须具有清晰的边界。然而,将某些不确定的对象强制分配到某个类中将降低聚类结果的结构和精度。三支聚类是一种重叠聚类,它采用核心域和边界域来表示每个类别,较好地处理了具有不确定性对象的聚类问题。提出了一种使用样本邻域将二支聚类转化为三支聚类的方法。该方法利用二支聚类的结果和每个类中元素的邻域是否完全包含在该类中来对集合进行收缩,同时利用不在该类中的元素的邻域是否与该类有交集来进行扩张。收缩的区域称为核心域,扩张域和核心域的差集称为边界域。在UCI数据集上的实验结果显示,该方法在提高聚类结果的结构和F1值方面有较好的效果。

关键词 三支聚类,邻域,K-means聚类,谱聚类
中图法分类号 TP391 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.01.009

Three-way Clustering Analysis Based on Dynamic Neighborhood

WANG Ping-xin[1,3] LIU Qiang[2] YANG Xi-bei[2] MI Ju-sheng[3]
(School of Science, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, China)[1]
(School of Computer Science, Jiangsu University of Science and Technology, Zhenjiang, Jiangsu 212003, China)[2]
(College of Mathematics and Information Science, Hebei Normal University, Shijiazhuang 050024, China)[3]

Abstract Most of the existing clustering methods are two-way clustering, which are based on the assumption that a cluster must be represented by a set with crisp boundary. However, assigning uncertain points into a cluster will reduce the accuracy of the method. Three-way clustering is an overlapping clustering which describes each cluster by core region and fringe region. This paper presented a strategy for converting a two-way cluster to three-way cluster using the neighborhood of the samples. In the proposed method, a two-way cluster is shrunk according to whether the neighborhood of sample are contained in this cluster and it is stretched according to whether the neighborhood of sample intersects with this cluster. The shrunk result is called core region and the difference between the shrunk result and stretched result is regarded as the fringe region. Experiment using the proposed method on UCI data sets shows that this strategy is effective in improving the structure and F1 values of clustering results.

Keywords Three-way clustering, Neighborhood, K-means clustering, Spectral clustering

1 引言

自Zadeh于1979年发表论文《Fuzzy sets and information granularity》[1]以来,研究人员对信息粒度化的思想产生了浓厚的兴趣。信息粒化是通过给定的粒化策略将面临的复杂数据粒化为信息粒的过程。根据不同的建模目标和用户需求,可以采用多种多样的粒化策略。聚类分析是一种常用的信息粒化方法,它是根据某一准则有机地将给定的数据集中的对象划分为若干个组或类的过程,通过聚类使得同一组内的数据对象具有较高的相似度,而不同组内的数据对象具有较高的相异度[2]。

聚类分析可以有效地发现事物之间的内在联系,描述隐含在数据集内部的结构特征。聚类分析已经在诸如图像处理[3]、网页搜索[4]和安全保障[5]等领域得到了成功的应用。特别是在生物技术领域,聚类分析通常被用来对动植物和基因进行分类,获取对种群固有结构的认识[6-7]。但是,目前在不确定性信息处理过程中,实际上大多数聚类方法是一种二支决策的结果,即决策一个对象要么属于某个类簇,要么不属

到稿日期:2017-03-03 返修日期:2017-05-07 本文受国家自然科学基金资助项目(61503160,61572242),江苏省高校自然科学基金(15KJB110004)资助。

王平心(1980—),男,博士,副教授,主要研究方向为粗糙集与粒计算,E-mail:pingxin_wang@hotmail.com(通信作者);刘 强(1989—),男,硕士生,主要研究方向为聚类分析,E-mail:qliu05@sina.com;杨习贝(1980—),男,博士,副教授,主要研究方向为粗糙集与粒计算,E-mail:zhenjiangyangxibei@163.com;米据生(1966—),男,博士,教授,主要研究方向为概念格和粒计算,E-mail:mijsh@263.net。

于某个类簇。然而在某些数据集上，传统的聚类方法并不能完全反映数据本身的结构特征。例如在图 1 中，若将论域中的点聚为两类，则可以明显地看到有两个数据集中的区域，对于元素 x_1 与 x_2，传统的二支决策聚类方法无论将它们分到哪一类，都不能显示出这些点的结构特征。

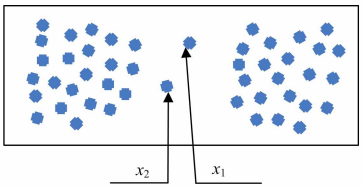

图 1　数据集示意图

Fig. 1　Schematic diagram of data set

2　相关工作

为了解决传统聚类方法存在的问题，很多学者对二支聚类算法进行了改进。Hoppner 等人[8] 提出了模糊聚类，采用模糊集来表示聚类的结果。Lingras[9-12] 提出了粗糙聚类方法，用粗糙集的正域、边界域和负域来表示聚类结果。Yao 等人[13] 采用区间集来表示一个类。上述方法都是对传统二支聚类方法的改进。

三支决策理论最早由 Yao 等人[14-16] 在决策粗糙集的基础上提出，其核心思想是将决策项拓展为正域决策 $POS(X)$、负域决策 $NEG(X)$ 和边界域决策 $BND(X)$，使它成为更符合人类认知的决策模式。三支决策的思想已广泛应用于医疗、教育以及管理等领域。Yu[17-20] 将三支决策的思想引入聚类中，提出了三支聚类方法，其思想是采用区间集表示一个类，其中区间集的下界称为类的核心区域(Core)，而位于区间上下界之间的对象组成类的边界区域(Fringe)，区间上界的补集称为类的琐碎域。

为了使聚类结果具有更好的结构特征，本文利用三支聚类的思想，在传统二支聚类的结果的基础上利用对象的 q 邻域介绍了一种三支聚类的算法，其主要思想是利用对二支聚类的结果进行收缩和扩张，从而得到三支聚类的核心域和边界域，同时其还考虑了该方法中 q 动态变化对聚类精度的影响。实验结果表明，该方法无论在聚类结果的结构上，还是在精度上都有很好的提升。以图 1 为例，利用传统硬聚类方法，x_1 与 x_2 只能属于某一个确定的类，假设聚类结果如图 2 所示，其中 x_1 与 x_2 分别被聚类到 C_2 与 C_1 中。本文的三支聚类方法的聚类结果如图 3 所示，其中 x_1 与 x_2 分别被聚类到 C_2 与 C_1 的边界域中，与传统的硬聚类方法相比，这样的聚类结果在结构上有明显的优势。

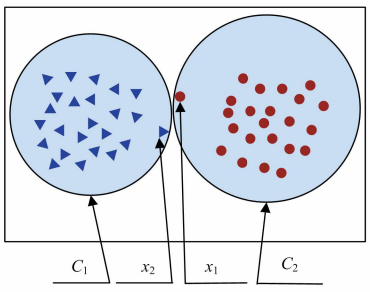

图 2　二支聚类的结果

Fig. 2　Clustering result by hard clustering

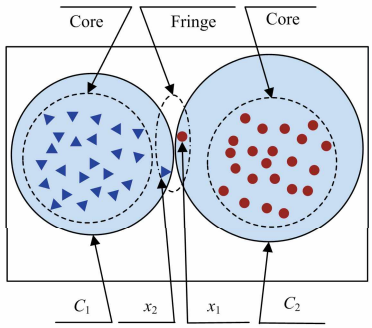

图 3　本文聚类方法的结果

Fig. 3　Clustering result by proposed method

3　三支聚类的相关概念

设给定一个数据 $U=\{x_1,x_2,\cdots,x_i,\cdots x_n\}$，传统的二支聚类方法是用一个集合表示一个类，即寻找一组集合 C_1，$C_2,\cdots,C_k$ 满足 $U=\bigcup_{i=1}^{k}C_i$ 且 $C_i\cap C_j=\varnothing(i,j=1,2,\cdots,k,i\neq j)$，其中 k 为聚类的个数。

基于三支决策的聚类思想是采用 3 个互不相交的集合表示一个类，即 C_i^P，C_i^B 与 C_i^N，分别称为类的核心域、边界域和琐碎域，其中：

$$C_i^P\cup C_i^B\cup C_i^N=U \tag{1}$$

若核心域中的元素确定属于某个类，则边界域中的元素可能属于也可能不属于这个类，而琐碎域中的元素肯定不属于这个类。由式(1)可知，可通过 C_i^P 和 C_i^B 来表示一个类。相反，若给定一组集合 C_i^P，$C_i^B(i=1,2,\cdots,k)$ 满足：

$$C_i^P\neq\varnothing,i=1,2,\cdots,k \tag{2}$$

$$\bigcup_{i=1}^{k}(C_i^P\cup C_i^B)=U \tag{3}$$

则称其为数据集 U 三支聚类。其中，式(2)要求正域非空，即每个类中至少有一个对象，而式(3)保证每个对象至少被分到一个类中。与传统硬聚类的结果 $C=\{C_1,C_2,\cdots,C_k\}$ 不同，三支聚类的结果应为：

$$TC=\{(C_1^P,C_1^B),(C_2^P,C_2^B),\cdots,(C_k^P,C_k^B)\} \tag{4}$$

显然在三支聚类中若有 $C_i^B=\varnothing(i=1,2,\cdots,k)$，则其变成了传统二支决策的聚类形式。因此，三支决策聚类形式是传统二支聚类方法的推广，也是针对一些不确定数据聚类问题的一种解决方案。针对目前知识体系下难以聚类的对象，我们无法确定其所属类别时，可将其归为某些类的边界域，等待新的信息以帮助进一步决策。

4　基于动态邻域的三支聚类算法

本节将介绍一种通过对二支聚类结果进行重构得到核心域和边界域的三支聚类方法。假设一个数据 $U=\{x_1,x_2,\cdots,x_i,\cdots,x_n\}$，$C=\{C_1,C_2,\cdots,C_k\}$ 是利用二支聚类算法对 U 进行聚类的结果。为了使聚类结果具有更好的结构，我们对每个类别 C_i 进行改进。以图 1 为例，事实上，若删除 x_1 与 x_2，则可以很容易地将图 1 聚成两个结构特征非常好的类，而无论将 x_1 或 x_2 放到哪一个类中都会降低这个类的紧致性。利用三支聚类的思想，我们将其放到 C_1 与 C_2 边界域中。而 x_1

与 x_2 这类点的显著特征就是在硬聚类的结果下，它们的 q 邻域(距离该点最近的 q 个点组成的集合)不完全包含于某个类中。将上述方法推广到一般的情况，可以得到如下的基于 q 邻域的三支聚类算法。

算法 1 基于 q 邻域的三支聚类算法

输入：数据集 $U=\{x_1,x_2,\cdots,x_i,\cdots,x_n\}$，参数 q，聚类数目 k

输出：$TC=\{(C_1^P,C_1^B),(C_2^P,C_2^B),\cdots,(C_k^P,C_k^B)\}$

Step1 利用硬聚类算法对 U 进行聚类，得到 $C=\{C_1,C_2,\cdots,C_k\}$；

Step2 对于每一类 C_i，任取 $x_j\notin C_i$，考虑 x_j 的 q 邻域 $Neig_q(x_j)$，若 $Neig_q(x_j)\cap C_i\neq\varnothing$，则 $x_j\in C_i^B$；

Step3 对于每一类 C_i，任取 $x_j\in C_i$，考虑 x_j 的 q 邻域 $Neig_q(x_j)$，若 $Neig_q(x_j)\subset C_i$，则 $x_j\in C_i^P$，否则 $x_j\in C_i^B$；

Step4 通过 Step2 和 Step3 得到 C_i^P 和 $C_i^B(i=1,2,\cdots,k)$，返回 $\{(C_1^P,C_1^B),(C_2^P,C_2^B),\cdots,(C_k^P,C_k^B)\}$。

在基于 q 邻域的三支聚类算法过程中，第一步是基于硬聚类的方法得到二支聚类的结果，而目前在二支聚类的算法中应用较为广泛的是 k-means 算法和谱聚类算法。下文将分别介绍基于 q 邻域的三支 k-means 聚类算法和三支谱聚类算法。

k-means 算法最早是由 MacQueen 于 1967 年提出的，是很典型的基于距离的聚类算法，采用距离作为相似性的评价指标，即认为两个对象的距离越近，其相似度就越大。该算法认为簇是由距离靠近的对象组成的，因此将获得的紧凑且独立的簇作为最终目标。该算法的第一步是随机地选取任意 k 个对象作为初始聚类的中心，初始地代表一个簇。该算法在每次迭代中根据数据集中剩余的每个对象与各个簇中心的距离将其重新赋给最近的簇。当考察完所有数据对象后，一次迭代运算完成，新的聚类中心被计算出来。若在一次迭代前后，新的聚类中心与原聚类中心相等或小于指定阈值，则算法结束。利用 k-means 聚类的结果，结合 q 邻域的三支聚类算法，可以得到基于 q 邻域的三支 k-means 聚类算法。

算法 2 基于 q 邻域的三支 k-means 聚类算法

输入：数据集 $U=\{x_1,x_2,\cdots,x_i,\cdots,x_n\}$，参数 q，聚类数目 k

输出：$TC=\{(C_1^P,C_1^B),(C_2^P,C_2^B),\cdots,(C_k^P,C_k^B)\}$

Step1 从 n 个初始数据中随机选取 k 个对象 $\mu_1,\mu_2,\cdots,\mu_k$ 作为聚类中心；

Step2 测量剩余的每个对象到每个聚类中心的距离，并把它归到最近的中心的类 $C_i=\{x_j\mid d(x_j,\mu_i)\leqslant d(x_j,\mu_l),l\neq i\}$；

Step3 重新计算已经得到的各个类的中心 $\mu_i=\frac{1}{|C_i|}\sum_{j\in C_i}x_j$，$(j=1,2,\cdots,k)$；

Step4 迭代 Step2 和 Step3，直至新的聚类中心与原聚类中心相等或小于指定阈值，算法结束，得到 $C=\{C_1,C_2,\cdots,C_k\}$；

Step5 对于每一类 C_i，任取 $x_j\notin C_i$，考虑 x_j 的 q 邻域 $Neig_q(x_j)$，若 $Neig_q(x_j)\cap C_i\neq\varnothing$，则 $x_j\in C_i^B$；

Step6 对于每一类 C_i，任取 $x_j\in C_i$，考虑 x_j 的 q 邻域 $Neig_q(x_j)$，若 $Neig_q(x_j)\subset C_i$，则 $x_j\in C_i^P$，否则 $x_j\in C_i^B$；

Step7 利用 Step5 和 Step6 得到 C_i^P 和 $C_i^B(i=1,2,\cdots,k)$，返回 $\{(C_1^P,C_1^B),(C_2^P,C_2^B),\cdots,(C_k^P,C_k^B)\}$。

谱聚类是一种基于图论的硬聚类方法，其基本思想是对利用样本数据的相似矩阵进行特征分解后得到的特征向量进行聚类。它将数据聚类问题看作一个无向图的多路划分问题。将数据点作为无向图 $G(V,E)$ 的顶点 V，边权重的集合 $E=\{W_{ij}\}$ 表示两点间的相似性，W 表示待聚类数据点间的相似性矩阵，将其看作该无向图的邻接矩阵，它包含了聚类所需要的所有信息。然后定义一个划分准则，优化该准则的目的是使同一类内的点具有较高的相似性，而不同类之间的点具有较低的相似性。由于求图划分判据的最优解是一个 NP 难问题，一个很好的求解方法是考虑问题的连续放松形式，这样便可将原问题转换成求图的 Laplacian 矩阵的谱分解，因此将这类方法统称为谱聚类。目前，已经出现了许多谱聚类模型和算法，如 Perona 和 Freeman[22] 提出的 PF 算法，Shi 和 Malik 提出的 SM[23] 算法，Scott 和 Longuet-Higgins[24] 提出的 SLH 算法，Ng，Jordan 和 Weiss[25] 提出的 NJW 算法等。利用谱聚类的结果，采用 q 邻域的三支聚类算法，可以得到基于 q 邻域的三支谱聚类算法。下面以经典 NJW 算法为基础，结合前文的三支聚类算法，给出基于 q 邻域的三支谱聚类算法的处理流程。

算法 3 基于 q 邻域的三支谱聚类算法(SPTHREE)

输入：数据集 $U=\{x_1,x_2,\cdots,x_i,\cdots,x_n\}$，尺度参数 σ，聚类数目 k

输出：$TC=\{(C_1^P,C_1^B),(C_2^P,C_2^B),\cdots,(C_k^P,C_k^B)\}$

Step1 将每个样本看作无向图的每个顶点来构建无向加权图，构造相似性矩阵 W，其中 $W_{ij}=\exp(-\frac{d^2(x_i,x_j)}{2\sigma^2})$，$i\neq j$，$W_{ii}=0$，$d(x_i,x_j)$ 为样本点 x_i 与 x_j 的欧氏距离；

Step2 计算相似度矩阵的度矩阵 $D_{ii}=\sum_j W_{ij}$；

Step3 构造拉普拉斯矩阵 $L=D^{-\frac{1}{2}}WD^{\frac{1}{2}}$；

Step4 计算矩阵 L 的前 k 个最大特征值及其对应的特征向量 $v_1,v_2,\cdots,v_k$，构造矩阵 $V=\{v_1,v_2,\cdots,v_k\}\in R^{n\times k}$；

Step5 规范化 V 的特征向量，得到矩阵 Y，其中 $Y_{ij}=\frac{V_{ij}}{\sqrt{\sum_j V_{ij}^2}}$；

Step6 将 Y 的每一行看作一个样本，然后使用 k-means 算法进行聚类；

Step7 若将 Y 的第 i 行数据归入第 j 类，则将原数据点 x_i 也划分到第 j 类，从而得到 U 的谱聚类 $C=\{C_1,C_2,\cdots,C_k\}$；

Step8 对于每一类 C_i，任取 $x_j\notin C_i$，考虑 x_j 的 q 邻域 $Neig_q(x_j)$，若 $Neig_q(x_j)\cap C_i\neq\varnothing$，则 $x_j\in C_i^B$；

Step9 对于每一类 C_i，任取 $x_j\in C_i$，考虑 x_j 的 q 邻域 $Neig_q(x_j)$，若 $Neig_q(x_j)\subset C_i$，则 $x_j\in C_i^P$，否则 $x_j\in C_i^B$；

Step10 利用 Step5 和 Step6 得到 C_i^P 和 $C_i^B(i=1,2,\cdots,k)$，返回 $\{(C_1^P,C_1^B),(C_2^P,C_2^B),\cdots,(C_k^P,C_k^B)\}$。

上文给出了基于 q 邻域的三支聚类算法，并在 k-means 算法和谱聚类的结果上进行了实现，基于 q 邻域的三支聚类算法和参数 q 有很大的关系，因此动态地考察聚类的结果和 q 的关系对选择合适的 q 有很大帮助。本文在实验算法中考虑了 q 的动态变化对聚类结果的影响。

5 实验结果

本节选用 5 组标准 UCI[26] 数据集对提出的算法进行测试，并通过相关实验，将 k-means 算法及谱聚类算法的结果与

文本提出的三支 k-means 算法及三支谱聚类的聚类结果进行对比分析。表 1 为实验中使用的 5 组测试数据集的基本信息。宏平均 F1 与微平均 F1 是评测系统分类性能的两个常用指标，其计算方法为：

$$Macro\ F1=\frac{2Macro\ P\times Macro\ R}{Macro\ P+Macro\ R}$$

$$Micro\ F1=\frac{2Micro\ P\times Micro\ R}{Micro\ P+Micro\ R}$$

其中，$Macro\ P=\frac{1}{k}\sum_{i=1}^{k}\frac{TP_{C_i}}{TP_{C_i}+FP_{C_i}}$，$Macro\ R=\frac{1}{k}\sum_{i=1}^{k}\frac{TP_{C_i}}{TP_{C_i}+FN_{C_i}}$，$Micro\ P=\frac{\frac{1}{k}\sum_{i=1}^{k}TP_{C_i}}{\frac{1}{k}\sum_{i=1}^{k}TP_{C_i}+\frac{1}{k}\sum_{i=1}^{k}FP_{C_i}}$，$Micro\ R=\frac{\frac{1}{k}\sum_{i=1}^{k}TP_{C_i}}{\frac{1}{k}\sum_{i=1}^{k}TP_{C_i}+\frac{1}{k}\sum_{i=1}^{k}FN_{C_i}}$。

其中，TP_{C_i} 表示 $Core(C_i)$ 中对象被正确分类为该类的数量，FP_{C_i} 表示非 $Core(C_i)$ 中对象被错误分类为该类的数量，FN_{C_i} 表示非 $Core(C_i)$ 中对象被正确分类为不属于该类的数量。

表 1　UCI 数据集描述

Table 1　Description of UCI data set

数据集	样本个数	样本维数	类别数
Banknote	1372	4	2
Hill Valley	1212	100	2
Ionosphere	351	34	2
SPECTF Heart	267	44	2
Vertebral Column	310	6	2

图 4—图 13 分别给出了 5 组数据集在 k-means 算法、谱聚类算法、三支 k-means 算法与三支谱聚类算法上的宏 F1 和微 F1 在实验中随参数 q 的动态变化的结果。从实验结果可以看出，无论是三支 k-means 算法还是三支谱聚类算法，其效果都优于原始硬聚类的算法，而且随着邻域 q 的不断变大，宏 F1 和微 F1 也在逐渐变大。

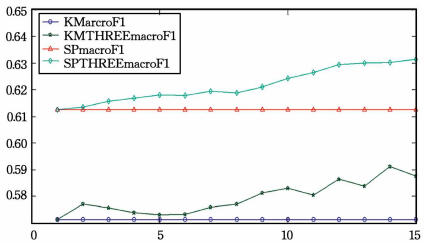

图 4　Banknote Authentication 集上的宏 F1 值

Fig. 4　Macro F1 value of Banknote Authentication

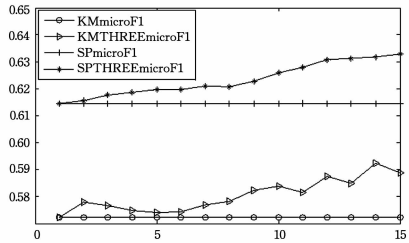

图 5　Banknote Authentication 集上的微 F1 值

Fig. 5　Micro F1 value of Banknote Authentication

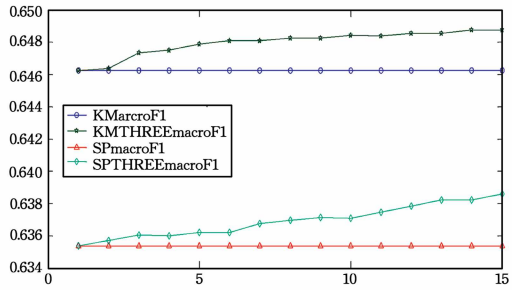

图 6　Hill Valley 集上的宏 F1 值

Fig. 6　Macro F1 value of Hill Valley

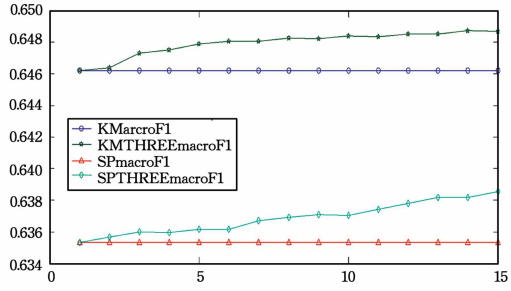

图 7　Hill Valley 集上的微 F1 值

Fig. 7　Micro F1 value of Hill Valley

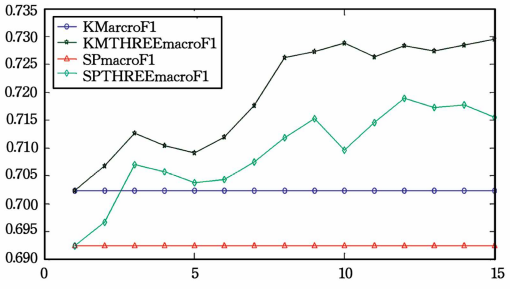

图 8　Ionosphere 集上的宏 F1 值

Fig. 8　Macro F1 value of Ionosphere

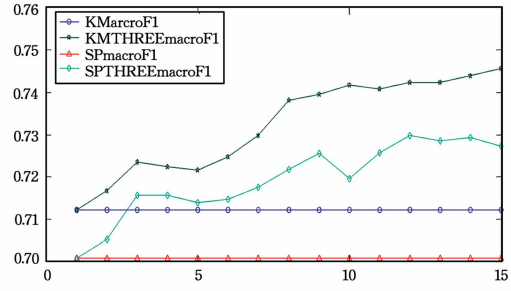

图 9　Ionosphere 集上的微 F1 值

Fig. 9　Micro F1 value of Ionosphere

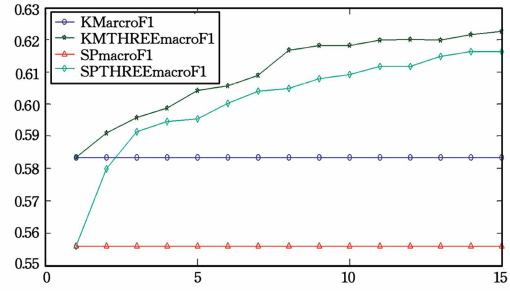

图 10　SPECTF Heart 集上的宏 F1 值

Fig. 10　Macro F1 value of SPECTF Heart

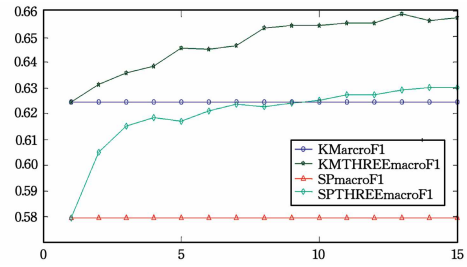

图 11 SPECTF Heart 集上的微 F1 值

Fig. 11 Micro F1 value of SPECTF Heart

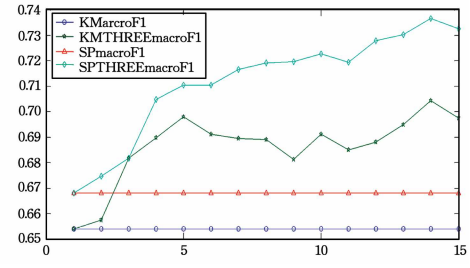

图 12 Vertebral Column 集上的宏 F1 值

Fig. 12 Macro F1 value of Vertebral Column

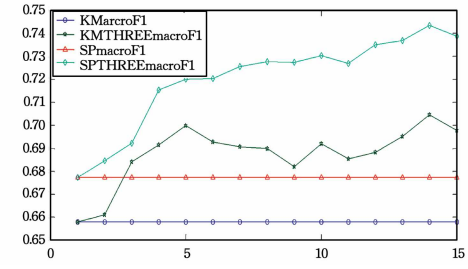

图 13 Vertebral Column 集上的微 F1 值

Fig. 13 Micro F1 value of Vertebral Column

结束语 考虑到在聚类过程中将不确定的对象强制分配到某个类中将降低聚类结果的结构和精度，文中通过三支聚类技术，利用二支聚类的结果和元素的邻域构建了一种将二支聚类转化为三支聚类的方法，并通过在 5 组 UCI 数据集上与传统方法的对比实验，验证了该方法在提高聚类结果的结构和精值方面有较好的效果。

在本文研究的基础上，下一步的研究可从以下两方面展开：1)本文考虑了在 k-means 聚类方法和谱聚类方法的结果上进行三支聚类的问题，还可以考虑在其他二支聚类的结果上进行三支聚类；2)利用动态邻域的思想，考虑基于动态邻域的特征选择问题。

参 考 文 献

[1] ZADEH L A. Fuzzy sets and information granulation. Advances in fuzzy set theory and applications[M]. Amsterdam: North-Holland Publishing, 1979: 35-48.

[2] SUN J G, LIU J, ZHAO L Y. Clustering algorithms research [J]. Journal of Software, 2008, 19(1): 48-61. (in Chinese)
孙吉贵，刘杰，赵连宇. 聚类算法研究[J]. 软件学报，2008，19(1)：48-61.

[3] ELALAMI M E. Supporting image retrieval framework with rule base system[J]. Knowledge-Based Systems, 2011, 24(2): 331-340.

[4] MARTIN-GUERRERO J D, PALOMARES A, BALAGUER-BALLESTER E, et al. Studying the feasibility of a recommender in a citizen web portal based on user modeling and clustering algorithms [J]. Expert Systems with Applications, 2006, 30(2): 299-312.

[5] KALYANI S, SWARUP K S. Particle swarm optimization based k-means clustering approach for security assessment in power systems [J]. Expert Systems with Applications, 2011, 38(9): 10839-10846.

[6] SHI J L, LUO Z G. Nonlinear dimensionality reduction of gene expression data for visualization and clustering analysis of cancer tissue samples [J]. Computers in Biology & Medicine, 2010, 40(8): 723-732.

[7] SEBISKVERADZE D, VRABIE V, GOBINET C, et al. Automation of an algorithm based on fuzzy clustering for analyzing tumoral heterogeneity in human skin carcinoma tissue sections [J]. Laboratory Investigation, 2011, 91(5): 799-811.

[8] HOPPNER F, KLAWONN F, KRUSE R, et al. Fuzzy cluster analysis: methods for classification, data analysis and image recognition [M]. Chichester: Wiley Press, 1999: 1-48.

[9] LINGRAS P. Rough K-Medoids clustering using Gas [C]// Proceedings of the 8th IEEE International Conference on Cognitive Informatics. Hong Kong: IEEE Press, 2009: 315-319.

[10] LINGRAS P, HOGO M, SNOREK M. Interval set clustering of web users using modified Kohonen self-organizing maps based on the properties of rough sets [J]. Web Intelligence and Agent Systems: An International Journal, 2004, 2(3): 217-230.

[11] LINGRAS P, HOGO M, SNOREK M, et al. Temporal analysis of clusters of supermarket customers: conventional versus interval set approach [J]. Information Sciences, 2005, 172(1/2): 215-240.

[12] LINGRAS P, WEST C. Interval set clustering of web users with rough K-Means [J]. Journal of Intelligent Information Systems, 2004, 23(1): 5-16.

[13] YAO Y Y, LINGRAS P, WANG R Z, et al. Interval Set Cluster Analysis: A Re-formulation [C]// International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing. Delhi: Springer, 2009: 398-405.

[14] YAO Y Y. Three-way decisions with probabilistic rough sets [J]. Information Sciences, 2010, 180(3): 341-353.

[15] YAO Y Y. The superiority of three-way decisions in probabilistic rough set models[J]. Information Sciences, 2011, 181(6): 1086-1096.

[16] YAO Y Y. An Outline of a Theory of Three-Way Decisions[C]// International Conference on Rough Sets and Current Trends in Computing. Berlin: Springer, 2012: 1-17.

[17] YU H, CHU S S, YANG D C. Autonomous knowledge-oriented clustering using decision-theoretic rough set theory [J]. Fundamenta Informaticae, 2012, 115(2-3): 141-156.

(下转第 89 页)

[10] BĚLOHLÁVEK R, SKLENÁŘV, ZACPAL J. Crisply generated fuzzy concepts [C]// International Conference on Formal Concept Analysis. Berlin Heidelberg: Springer, 2005: 269-284.

[11] KUMAR C A, SRINIVAS S. Concept lattice reduction using fuzzy K-means clustering[J]. Expert systems with applications, 2010, 37(3): 2696-2704.

[12] WU W Z, LEUNG Y, MI J S. Granular Computing and Knowledge Reduction in Formal Contexts [J]. IEEE Transactions on Knowledge and Data Engineering, 2009, 21(10): 1461-1474.

[13] LI J H, MEI C L, LV Y J. Knowledge Reduction in Decision Formal Contexts [J]. Knowledge-Based Systems, 2011, 24(5): 709-715.

[14] LI J H, MEI C L, LV Y J. Incomplete Decision Contexts: Approximate Concept Construction, Rule Acquisition and Knowledge Reduction [J]. International Journal of Approximate Reasoning, 2013, 54(1): 149-165.

[15] SHAO M W, LEUNG Y, WU W Z. Rule Acquisition and Complexity Reduction in Formal Decision Contexts[J]. International Journal of Approximate Reasoning, 2014, 55(1): 259-274.

[16] LI L J, MI J S, XIE B. Attribute Reduction Based on Maximal Rules in Decision Formal Context[J]. International Journal of Computational Intelligence Systems, 2014, 7(6): 1044-1053.

[17] ZHANG W X, WEI L, QI J J. Attribute Reduction Theory and Approach to Concept Lattice[J]. Science in China Series E, 2005, 35(6): 628-639. (in Chinese)
张文修，魏玲，祁建军. 概念格的属性约简理论与方法[J]. 中国科学：E 辑，2005，35(6)：628-639.

[18] MI J S, LEUNG Y, WU W Z. Approaches to Attribute Reduction in Concept Lattices Induced by Axialities[J]. Knowledge-Based Systems, 2010, 23(6): 504-511.

[19] SHAO M W, LEUNG Y, WANG X Z, et al. Granular Reducts of Formal Fuzzy Contexts [J]. Knowledge-Based Systems, 2016, 114: 156-166.

[20] LI M Z, WANG G Y. Approximate Concept Construction With Three-Way Decisions and Attribute Reduction in Incomplete Contexts [J]. Knowledge-Based Systems, 2016, 91: 165-178.

[21] WANG J H, LIANG J Y, QIAN Y H. A Heuristic Method to Attribute Reduction for Concept Lattice [C]// International Conference on Machine Learning and Cybernetics. New York: IEEE Press, 2010, 1: 483-487.

[22] LV Y J, LI J H. Heuristic Algorithms for Attribute Reduction on Concept Lattice[J]. Computer Engineering and Applications, 2009, 45(2): 154-157. (in Chinese)
吕跃进，李金海. 概念格属性约简的启发式算法[J]. 计算机工程与应用，2009，45(2)：154-157.

[23] LIANG J Y, XU Z B. Uncertainty Measures of Roughness of Knowledge and Rough Sets in Incomplete Information Systems [C]// Proceedings of the 3rd World Congress on Intelligent Control and Automation, 2000. New York: IEEE Press, 2000: 2526-2529.

[24] LIANG J Y, XU Z B. The Algorithm on Knowledge Reduction in Incomplete Information Systems [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(1): 95-103.

[25] HUANG B, ZHOU X Z, SHI Y C. Entropy of Knowledge and Rough Set Based on General Binary Relation [J]. Systems Engineering-Theory & Practice, 2004, 24(1): 93-96. (in Chinese)
黄兵，周献中，史迎春. 基于一般二元关系的知识粗糙熵与粗集粗糙熵[J]. 系统工程理论与实践，2004，24(1)：93-96.

[26] GANTER B, WILLE R. Formal Concept Analysis: Mathematical Foundations [M]. Berlin, Germany: Springer, 1999.

[27] YAO Y Y, ZHAO Y. Discernibility matrix simplification for constructing attribute reducts[J]. Information Sciences, 2009, 179(7): 867-882.

[28] LICHMAN M. UCI machine learning repository [EB/OL]. http://archive.ics.uci.edu/ml.

(上接第 66 页)

[18] YU H, LIU Z G, WANG G Y. An automatic method to determine the number of clusters using decision-theoretic rough set [J]. International Journal of Approximate Reasoning, 2014, 55(1): 101-115.

[19] YU H, ZHANG C, WANG G Y. A tree-based incremental overlapping clustering method using the three-way decision theory [J]. Knowledge-Based Systems, 2016, 91(C): 189-203.

[20] YU H, JIAO P, YAO Y Y, et al. Detecting and refining overlapping regions in complex networks with three-way decisions [J]. Information Sciences, 2016, 373(1): 21-41.

[21] MACQUEEN J B. Some methods for classification and analysis of multivariate observations [C]// Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1967: 281-297.

[22] PERONA P, FREEMAN W T. A Factorization Approach to Grouping [C]// European Conference on Computer Vision. Berlin: Springer, 1998: 655-670.

[23] SHI J, MALIK J. Normalized cuts and image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(8): 888-905.

[24] SCOTT G L, LONGUET-HIGGINS H C. Feature grouping by relocalisation of eigenvectors of proximity matrix [C]// Proceedings of British Machine Vision Conference. Oxford: BMVA Press, 1990: 103-108.

[25] NG A, JORDAN M, WEISS Y. On spectral clustering: analysis and an algorithm[C]// International Conference on Neural Information Processing Systems: Natural and Synthetic. Shanghai: MIT Press, 2001: 849-856.

[26] UCI machine Learning Repository [OL]. http://www.ics.uci.edu/mlearn/MLRepository.html.