

# 一种基于投票的三支决策聚类集成方法

胡凌超, 于 洪

(重庆邮电大学 计算智能重庆市重点实验室 重庆 400065)

E-mail: yuhong@cqupt.edu.cn

**摘 要:** 聚类集成通过综合多个聚类结果得到最终的类簇划分,在保证与每个被集成的聚类结果尽量一致的前提下,提高了聚类算法的稳定性。但是目前少有专门针对二支决策聚类结果集成为三支决策聚类结果的集成方法,而三支决策以其更符合人类认知与决策模式的优势正被学术界广泛研究。为此,提出一种新的基于三支决策的聚类集成方法,并用区间集的形式表示最终聚类结果,即用区间集的下界、边界与上界分别表示聚类结果中类簇的正域、边界域和负域。首先,通过已有的聚类算法得到相同数据集不同的二支决策聚类结果,然后通过聚类成员类簇标签匹配以及求交集的方式得到一部分数据的类簇归属,最后通过基于投票的三支决策方法,对剩余部分数据进行类簇划分。实验结果表明文中提出的基于三支决策的聚类集成方法是有效的。

**关键词:** 聚类集成; 三支决策; 区间集; 投票

中图分类号: TP18

文献标识码: A

文章编号: 1000-1220(2016)08-1741-05

## Voting Cluster Ensemble Approach Based on Three-way Decisions

HU Ling-chao, YU Hong

(Chongqing Key Lab of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

**Abstract:** Cluster ensemble can combine the outcomes of several clusterings to a single clustering that agrees as much as possible with the input clusterings and improves the stability of the clustering algorithm. However, little attention has been paid to the approaches to deal with consolidating the outcomes of hard clustering, which is a result of two-way clustering, into a soft clustering, which is a result of three-way clustering. Since the three-way decision theory is widely studied recently with its advantages of human cognition and decision-making, the main objective of this paper is to propose a transfer approach based on the framework of cluster ensemble. Here, an interval set is used here to represent a cluster, which is described by three regions according to the lower and upper bound of the cluster. First, the proposed approach gets different hard partitions through ensemble members which are the existing clustering algorithms. Then, it decides some objects definitely to the corresponding clusters by matching the clusters' tag and intersection operations. Finally, it decides the rest objects by three-way decisions based on voting. The preliminary experimental results show that the proposed approach is effective.

**Key words:** cluster ensemble; three-way decisions; interval sets; voting

## 1 引言

聚类一直是机器学习、数据挖掘、模式识别和图像分析等领域一个比较活跃而且极具挑战性的研究方向。

虽然目前已经有许多聚类算法,但没有一种单一的聚类方法能够识别出任意形态的数据结构分布<sup>[2,3]</sup>。加上现有的聚类算法存在着以下局限性<sup>[4]</sup>:大多数聚类算法难于判断数据集的真实类簇个数;聚类结果很大程度上取决于参数及初始化;对同一数据集不同的聚类算法可能产生不同的结果。在这种情况下,综合多个聚类结果得到最终类簇的聚类集成方法被提出并得到了广泛的关注,且表现出了比单个聚类算法更高的鲁棒性与稳定性<sup>[1]</sup>。

聚类集成主要包含两个过程:生成过程与一致性划分。本文方法主要是解决已生成的聚类结果的一致性划分问题。目前为止,已有很多学者对一致性划分问题做了研究,如: Fred<sup>[5]</sup>提

出了基于投票的 $k$ 均值聚类集成算法;Strehl和Ghosh<sup>[6]</sup>利用图理论知识,提出了三个基于超图的方法:HGPA、MCLA和CSPA;Wang等<sup>[7]</sup>考虑了类簇大小、样品维度和密度等因素,提出了概率积累的聚类集成方法;周等<sup>[8]</sup>利用矩阵谱分析理论提出了一种基于谱聚类的聚类集成算法等。

尽管聚类集成方法有很多,但是对于将多个硬聚类结果或者软聚类结果集成为一个软聚类结果的研究还相对较少。一些学者通过模糊集理论知识对集成获得软聚类结果的情况进行了研究,如: Punera等<sup>[9]</sup>将文献[6]中CSPA、HGPA和MCLA三种硬聚类方法改进为软聚类方法,从而得到一个软聚类结果;Sevillano<sup>[10]</sup>等使用隶属度矩阵表示聚类结果,其所提出的一致性函数可以将多个软聚类结果集成一个最终的软聚类结果;Avogadri等<sup>[11]</sup>提出一种非监督模糊聚类集成算法,使用模糊 $k$ 均值算法作为基聚类器,可以获得软聚类结果或硬聚类结果等。这些方法使用了模糊集理论知识中的隶属

收稿日期: 2015-06-04 收修改稿日期: 2015-07-30 基金项目: 国家自然科学基金项目(61379114, 61272060) 资助; 重庆市自然科学基金项目(cstc2013jcyjA40063) 资助。 作者简介: 胡凌超,男,1991年生,硕士研究生,研究方向为数据挖掘、聚类分析;于 洪,女,1972年生,博士,教授,研究方向为粗糙集、三支决策、智能信息处理、Web 智能、数据挖掘等。



度来表示数据对象隶属类的程度。

另外一方面,硬聚类结果是典型的二支决策的结果,即:一个对象要么属于这个类簇,要么不属于这个类簇,而类簇的表示通常采用集合的方式.对于软聚类结果,目前常用的表示方式也是基于集合的,只不过是一个对象可能属于多个类簇集合而已.但是,这种表示不便直接地观察哪些对象是属于多个类簇的. Lingras 和 Yan<sup>[12]</sup>认为使用区间集可以更好地表示聚类,尤其在数据分类时能够更好描述其模糊性.因此,我们采用了区间集的形式来描述类簇,即,通过区间集的上下界将类簇分为了三个域:正域、边界域和负域<sup>[16]</sup>.在现有信息可以确定数据对象类簇归属的情况下,将该数据对象划分在其所属类簇的正域部分,即该类簇的下界;在现有信息不能确定数据对象是否属于某个类时,则将该数据对象划分在该类的边界域部分,即,该类簇上界与下界的差集;而明确不属于某类的数据对象则划分在该类簇的负域部分,其实也就是其他类簇的正域或者边界域部分.

因此,本文针对二支决策聚类结果如何转换为一个三支决策聚类结果的问题,基于集成聚类的框架通过投票方式给出了一种解决方案.具体说来就是:在对二支决策聚类结果集成的过程中,我们首先对所有聚类成员中的类簇做标签匹配;然后,对所有具有相同标签的类簇进行交集运算,交集部分的数据可认为是类簇归属已经明确了,因为这部分数据在每个聚类成员中所归属的类簇的标签都是一致的,即,都划分在了同一个类中,所以先将这部分数据划分在相应标签的类簇的正域部分;最后,通过构建矩阵记录剩余部分数据在各个类簇中的归属情况,根据本文提出的基于投票的三支决策聚类方法,判断数据对象划分在类簇的正域部分或者边界域部分.由此,便可得到全部数据对象的类簇归属,将二支决策聚类结果集成为一个三支决策聚类结果.

本文剩下的部分组织如下:第2部分介绍了聚类集成、区间集的一些基本理论以及三支决策聚类结果用区间集表示的方法.第3部分介绍了本文提出的基于投票的三支决策聚类集成方法.第4部分提供了实验结果及其分析.第5部分对全文进行了总结,并指出未来的研究方向.

## 2 基本理论

### 2.1 聚类集成

一般说来,聚类集成过程包括两个阶段:生成过程(Generation Process)和一致性划分(Consensus Fuction).

假设目前有  $N$  个数据对象构成的数据集:  $U = \{x_1, \dots, x_n, \dots, x_N\}$ , 其中各个数据  $x_n$  拥有  $D$  维属性,表示为  $x_n = (x_n^1, \dots, x_n^d, \dots, x_n^D)$ ,  $x_n^d$  表示数据  $x_n$  的第  $d$  维属性( $n \in \{1, \dots, N\}$ ,  $d \in \{1, \dots, D\}$ ).

对数据集做聚类集成,首先采用生成过程:选取  $H$  个聚类算法  $\Phi = \{\Phi_1, \dots, \Phi_h, \dots, \Phi_H\}$ , 其中  $\Phi_h$  为第  $h$  个聚类算法,或者是选取同一个聚类算法进行  $H$  次不同的初始化,然后对数据集  $U$  进行聚类,得到  $H$  个聚类成员,即  $P = \{P_1, \dots, P_h, \dots, P_H\}$ , 其中  $P_h = \{C_h^1, \dots, C_h^k, \dots, C_h^K\}$ ,  $C_h^k$  表示第  $h$  个聚类成员的第  $k$  个类簇,  $1 \leq k \leq K$ ,  $K$  为类簇的个数(除此之外,聚类成员还可通过选取不同的数据子集及选择不同的特征子集投影到数据子空间的方式获得);然后进行一致性划分:根

据一致性函数将这  $H$  个聚类成员进行融合,得到一个最终的聚类结果  $C = \{C^1, \dots, C^k, \dots, C^K\}$ .

### 2.2 区间集简介及三支决策聚类结果的区间集表示

Yao 和 Lingras 在文献<sup>[13]</sup>中引入 Interval Set,使用区间集来代替单一集合的表达方式.

区间集正好体现了三支决策的思想,当使用 Interval Set 表示聚类结果时,类簇使用一对集合来表示.类簇被分为三个部分:正域部分(POS)、边界域部分(BND)和负域部分(NEG).当数据对象无法确定是否属于该类簇时,就将其划分到边界域部分.位于边界域部分的数据对象可能属于多个类簇,从而能够描述其模糊性.

因此我们采用区间集的形式表示一个类簇:

$$C^k = [\underline{C}^k, \overline{C}^k]$$

其中  $\underline{C}^k$  表示聚类结果中第  $k$  个类簇的下边界集,  $\overline{C}^k$  表示该类簇的上边界集,且  $\underline{C}^k \subseteq \overline{C}^k$ . 而三个域的表现形式则为:

$$\begin{aligned} POS(C^k) &= \underline{C}^k \\ BND(C^k) &= \overline{C}^k - \underline{C}^k \\ NEG(C^k) &= U - \overline{C}^k \end{aligned}$$

这样,最终的聚类结果  $C$  则表示为:

$$C = \{[\underline{C}^1, \overline{C}^1], \dots, [\underline{C}^k, \overline{C}^k], \dots, [\underline{C}^K, \overline{C}^K]\}$$

其中类簇具备如下性质:

$$(i) \underline{C}^k \neq \emptyset, 1 \leq k \leq K; (ii) \bigcup \overline{C}^k = U$$

性质(i)表示每一个类簇不能为空集;性质(ii)表示每一个数据必须至少属于一个类簇的上确界,即代表每个数据都被合理地做了聚类决策,这样便保证了聚类结果的意义.

## 3 基于投票的三支决策聚类集成方法

### 3.1 基于投票的三支决策聚类集成方法框架图

图1为本文提出的基于投票的三支决策聚类集成方法的框架图.

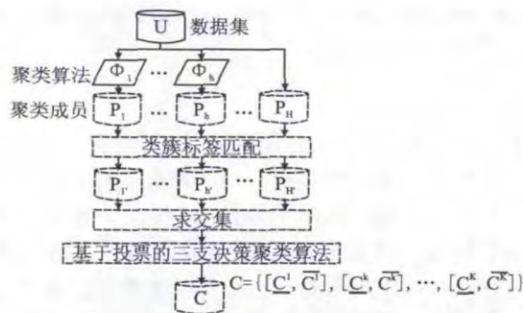


图1 基于投票的三支决策聚类集成方法框架图

Fig. 1 Framework of voting clustering ensemble model based on three-way decisions

### 3.2 聚类成员的生成

在以往的聚类集成算法中,有的聚类成员为硬聚类结果,有的为软聚类结果,也有算法可以同时处理包含了硬聚类与软聚类两种结果的聚类成员.而本文算法所处理的聚类成员只限于硬聚类结果,即只对二支决策聚类结果进行集成,在一定程度上降低了集成的复杂性.

聚类成员可以是通过已有的聚类算法获得,也可以是已



知的对于数据集的二支决策的划分结果. 这样便扩大了本文算法框架能够应用的范围,即,当我们拥有聚类算法时,可以对数据集进行聚类产生聚类成员,不拥有聚类算法而拥有已知的数据集的划分结果时,也可以采用本文模型进行集成,而二者结合当然也同样适用.

同时,考虑到如果不同的聚类成员中的类簇个数不同,可能出现两种情况:1) 部分聚类成员中的类簇个数小于最终结果的类簇个数,在进行后序的三支决策方法时可能会降低正确率;2) 部分成员的类簇个数大于最终结果的类簇个数,则在处理过程中,增加了复杂性. 本文算法为了简便,设定了前提条件为每个聚类成员中类簇的个数是相同的. 而如何将算法应用在不同类簇个数的情况是后期进一步研究的方向.

### 3.3 算法中的三支决策方法

#### 3.3.1 聚类成员类簇标签匹配

聚类结果中的每个类簇,都可以用符号来标记,如: $r, s, t$ 等,这些符号本文称之为“标签”(其实也就是每个类的类号),并用符号  $L(x)$  表示类簇  $x$  的标签.

由于目前还未对不同的聚类成员中的类簇做标签匹配,数据在不同的聚类成员中得到的划分,可能会呈现这样的情况,如: $[1, 2, 2, 1, 1, 3, 3, 3], [2, 3, 3, 2, 2, 1, 1, 1]$ ,此时,它们的表达方式虽然不同,但是却表达着同一个聚类结果. Tang 等人在文献[15]中提到:一般来说,有对应关系的类簇所覆盖的相同对象个数应该是最大的. 因此,根据这一启发式我们可以对类簇进行标签匹配,如:聚类成员  $P_i = \{C_i^1, \dots, C_i^k, \dots, C_i^K\}$ ,  $P_j = \{C_j^1, \dots, C_j^k, \dots, C_j^K\}$ ,  $1 \leq i, j \leq H$ . 首先,将两个划分中每一个类簇  $C_i^{k_1}, C_j^{k_2}$  ( $1 \leq k_1, k_2 \leq K$ ) 所覆盖的相同对象的个数记录在  $K \times K$  的 OVERLAP 矩阵中,然后选择其中覆盖相同对象个数最大的类簇标签建立对应关系,并将结果从 OVERLAP 矩阵中移除. 重复以上过程,直到所有的类簇标签都建立了对应关系为止. 当存在两个及以上类簇标签对应时,则随机选取某个类簇标签作为匹配<sup>[15]</sup>.

为了方便,我们设定所有的聚类成员以第一个聚类成员  $P_1$  中的类簇划分作为参照,进行标签匹配,并将聚类成员中被匹配类簇的标签改为在  $P_1$  中所对应的类簇的标签,使得  $L(P_1^k) = \dots = L(P_h^k) = \dots = L(P_H^k)$ ,  $1 \leq k \leq K$ .

类簇标签匹配算法的伪代码为:

算法:类簇标签匹配

输入:未经标签匹配的  $H$  个聚类结果  $P = \{P_1, \dots, P_h, \dots, P_H\}$

输出:经标签匹配之后的  $H$  个聚类结果  $P = \{P_1, \dots, P_h, \dots, P_H\}$

步骤:

1. 选择  $P$  中的  $P_1 = \{P_1^1, \dots, P_1^k, \dots, P_1^K\}$  作为参照划分;
2. For ( $j=2; j \leq H; j++$ )
3. {
4. 将  $P_j = \{P_j^1, \dots, P_j^k, \dots, P_j^K\}$  中的  $K$  个类簇与  $P_1$  中的类簇构建  $K \times K$  重叠度矩阵,其中行号代表  $P_1$  中的类簇标签,列号代表  $P_j$  中的类簇标签; // 重叠度矩阵即类簇之间相同数据个数构成的矩阵
5. For ( $k=1; k \leq K; k++$ )
6. | 选择矩阵中最大数字,并从矩阵中取出数字所对应的行与列,列号 = 行号;|
7. |
8. 输出结果.

#### 3.3.2 相同标签类簇求交集

在所有的聚类成员类簇标签匹配之后,求取所有聚类成员中相同标签的类簇的交集  $P_1^k \cap \dots \cap P_h^k \cap \dots \cap P_H^k$ , 其中  $1 \leq k \leq K$ . 由于在每个聚类成员中,这部分数据都划分在了同一个类簇中,我们认为这部分数据是明确归属于其所属类簇的,即,这部分数据应当划分在该类簇的正域 POS 中,假设这个类为第  $k$  类,则划分在  $POS(C^k)$ , 也就是  $C^k$  中.

#### 3.3.3 基于投票的三支决策聚类算法

这一小节,我们将对剩余部分数据采用基于投票的三支决策聚类算法进行类簇归属的划分.

划分的过程类似于投票选举,每一个聚类成员可以看成是一个投票者,它们将对每一个剩余的数据投票一次且仅有一次,而对每个数据的投票过程中,都有  $K$  种票供聚类成员选择,每个聚类成员只能选择其中的一种票投给一个数据点. 也就是说:每个数据点在投票结束时都会获得来自  $H$  个聚类成员每位成员的一张票,总共  $H$  张票,即:它们在  $H$  个聚类成员中所归属的类簇的标签. 这样我们就可以构建一个矩阵来记录每个剩余数据点获得的  $K$  类票的张数;假设剩余数据量的大小为  $N_2$ ,用  $N_2 \times K$  的矩阵,记录投票情况. 然后,我们将根据投票的情况,对数据的归属做三支决策.

对于剩余的每个数据点,我们按照如下的三支决策规则,依次对其进行类簇归属的划分:

**第1步.** 判断该数据点在所有类( $K$ 类)的  $H$  个得票数中,是否存在大于等于  $H/2$  的. 如果存在,则对这个数据做正域决策,即,将该数据点放置在该类的正域(POS)部分;如果不存在,则进行第2步.

**第2步.** 判断是否存在票数大于等于  $H/4$  的类,如果存在,则对该点做边界域决策,即,将该数据点放置在这些类的边界域(BND)部分;如果不存在,则进行第3步.

**第3步.** 找出该点所得票数中大于0的类,将数据点划分在这些类的边界域部分.

这样我们便得到了所有数据点的三支决策聚类结果.

表1 投票矩阵  $N_2 \times K$

Table 1  $N_2 \times K$  matrix

数据点 \ 簇	C1	C2	C3	C4	C5	C6
X1	12	0	1	0	7	0
X2	1	2	3	4	0	10
X3	7	6	2	3	1	1
X4	0	9	5	5	0	1
X5	4	0	4	4	4	4

下面,我们举个例子进一步解释基于投票的三支决策算法:

假设我们当前剩余  $N_2 = 5$  个数据点,有20个聚类成员对它们进行投票,类簇的个数  $K = 6$ . 此时投票矩阵  $N_2 \times K$  如表1所示.

首先对第一个数据点做判断:它在第1类的票数为12,大于  $20/2$ ,此时,直接将该数据点划分在第1类的正域部分  $POS(C^1)$  中;

然后是第二个点,票数中第6类为10,等于  $20/2$ ,直接将点划分在第6类的正域部分  $POS(C^6)$  中;



第三,四个点中都不存在大于等于 10 的票数,但是存在大于等于 20/4 的. 其中第三个点在第 1,2 类中分别为 7,6,所以将该点划分在  $BND(C^1)$ ,  $BND(C^2)$  中;而第四个点划分在  $BND(C^2)$ ,  $BND(C^3)$ ,  $BND(C^4)$  中.

第五个点的票数中既不存在大于等于 20/2 的,也不存在大于等于 20/4 的,所以判断其中票数大于 0 的类,为:1,3,4,5,6,所以将其划分在  $BND(C^1)$ ,  $BND(C^3)$ ,  $BND(C^4)$ ,  $BND(C^5)$ ,  $BND(C^6)$  中.

4 实验分析

实验中,本文聚类成员的生成采用  $k$  均值算法,且只采用了  $k$  均值算法获得的二支决策聚类结果作为聚类成员. 本文的正确率是基于三支决策聚类结果中的正域部分给出的.

实验 1. 首先用 2 维人工数据集来展示本文算法的效果,

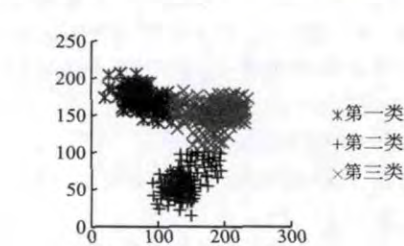


图 2 原数据集分类情况  
Fig. 2 Initial dataset

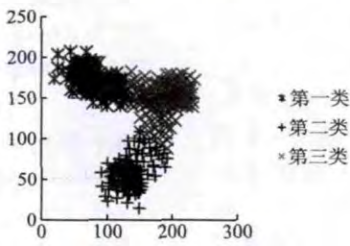


图 3  $k$  均值算法聚类结果  
Fig. 3 Clustering results of  $k$ -means

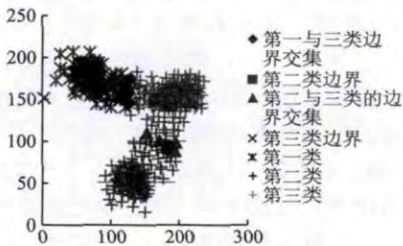


图 4 本文算法聚类结果  
Fig. 4 Clustering results of the proposed approach

实验 2. 此外,为了进一步验证本文提出算法的性能,我们在一些真实数据集上进行了实验,选取的真实数据集均来自 UCI 机器学习数据库,表 2 给出了这些数据集的信息.

表 2 实验中使用的数据集  
Table 2 Data sets used in the experiment

Data sets	Number of samples	Number of features	Number of classes
Iris	150	4	3
Wine	178	13	3
Ionosphere	351	34	2
Liver Disorders	345	6	2
Image Segmentation	2310	19	7
Waveform	5000	21	3
Page Blocks	5473	10	5

在对每一组数据的实验中,本文算法以及  $k$  均值算法都分别运行 10 次获得准确率,然后计算它们准确率的平均值以及在 10 次实验中的波动情况. 而作为与本文算法对比的另一个算法<sup>[14]</sup>的实验结果,则是直接从原文中获取了它们实验结果的相关部分.

从表 3 可以看出,本文算法稳定性比较好. 相较于文献<sup>[14]</sup>所提出的集成算法,本文算法在 Iris 数据集上的准确率高于它,在 Image Segmentation 数据集上的准确率低于它,而在其他数据集上则与之效果基本持平. 而对比  $k$  均值算法,本文所提出的算法在 Wine 和 Image Segmentation 两个数据集上的准确率高于  $k$  均值算法,在 Ionosphere 数据集上也略高

该数据集包含 480 个数据对象,类簇个数  $K$  为 3. 在实验中,参数  $H$  设置为 10,即产生 10 个聚类成员.  $k$  均值的聚类次数阈值为 50. 同时,为了更好地体现本文算法的优越性,我们同时采用了单个  $k$  均值算法对数据集进行聚类,并将原数据集及两种算法得到的结果通过图的方式进行展示:图 2 为原数据集分类情况的展示,图 3 为  $k$  均值算法的聚类结果,图 4 为本文算法的聚类结果展示.

由图 2 与图 4 对比可以看到,本文算法基本保留了原有数据集的分布,即:绝大部分数据都与数据的原划分是一致的,说明本文算法能够得到较好的聚类效果. 由图 3 与图 4 对比可以看到,本文算法在类簇的交界处,即,数据的类簇归属不是那么明显的地方或将数据单独划分在了类簇的边界域,或将它们同时划分在了两个类的边界域中,这样处理比  $k$  均值算法一味地将数据非此即彼地划分显得更为合理.

于它,而在其他的数据集上的效果与  $k$  均值算法持平. 综上所述,本文所提出的算法能够较好地得到聚类结果.

表 3 UCI 数据集上的实验结果  
Table 3 Accuracy of different algorithms on UCI data sets

Data sets	the proposed approach	the algorithm in [14]	the k-means algorithm
Iris	0.89 ± 0.00	0.87 ± 0.04	0.89 ± 0.00
Wine	0.70 ± 0.00	0.70 ± 0.01	0.66 ± 0.03
Ionosphere	0.71 ± 0.00	0.70 ± 0.03	0.70 ± 0.01
Liver Disorders	0.55 ± 0.00	0.55 ± 0.00	0.55 ± 0.00
Image Segmentation	0.55 ± 0.06	0.60 ± 0.04	0.52 ± 0.06
Waveform	0.39 ± 0.00	0.39 ± 0.00	0.39 ± 0.00
Page Blocks	0.73 ± 0.00	0.73 ± 0.00	0.73 ± 0.00

5 结 论

本文提出了一种基于投票的三支决策聚类集成方法. 方法只针对数据集的二支决策聚类结果构成的聚类成员,采用结合了类簇标签匹配求交集对一部分数据做正域决策,基于投票法的三支决策聚类规则对剩余部分数据聚类的集成方法,得到数据集的三支决策聚类结果. 实验结果表明本文提出的聚类集成算法是有效的. 未来的工作是进一步细化三支决策规则,得到准确率更高的结果.

References:

[1] Yu Hong, Liu Zhan-guo, Wang Guo-yin. 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.
- [2] Naldi M C, Carvalho A, Campello R J G B. Cluster ensemble selection based on relative validity indexes[J]. *Data Mining and Knowledge Discovery*, 2013, 27(2): 259-289.
- [3] Minaei Bidgoli Behrouz, Topchy Alexander, Punch William F. Ensembles of partitions via data resampling[C]. *Proceedings of International Conference on Information Technology: Coding and Computing*, Washington, 2004, 2: 188-192.
- [4] Dietterich T G. Ensemble methods in machine learning[M]. *Lecture Notes in Computer Science*, Springer, 2000: 1-15.
- [5] Fred A. Finding consistent clusters in data partitions[M]. *Multiple-Classifer Systems*, Berlin, Springer Berlin Heidelberg, 2001, 2096: 309-318.
- [6] Strehl A, Ghosh J. Cluster ensembles—a knowledge reuse framework for combining multiple partitions[J]. *Journal of Machine Learning Research*, 2002, 3(12): 583-617.
- [7] Wang Xi, Yang Chun-yu, Zhou Jie. Clustering aggregation by probability accumulation[J]. *Pattern Recognition*, 2009, 42(5): 668-675.
- [8] Zhou Lin, Ping Xi-jian, Xu Sen, et al. Clustering ensemble algorithm based on spectral clustering[J]. *Acta Automatica Sinica*, 2012, 38(8): 1335-1342.
- [9] Punera K, Ghosh J. Consensus-based ensembles of soft clusterings[J]. *Applied Artificial Intelligence*, 2008, 22(7-8): 780-810.
- [10] Xavier S, Alians F, Socoro J C. Positional and confidence voting-based consensus functions for fuzzy cluster ensembles[J]. *Fuzzy Sets and Systems*, 2012, 193: 1-32.
- [11] Avogadri R, Valentini G. Ensemble clustering with a fuzzy approach[M]. *Supervised and Unsupervised Ensemble Methods and their Applications*, Berlin: Springer Berlin Heidelberg, 2008.
- [12] Lingras P, Yan R. Interval clustering using fuzzy and rough set theory[C]. *IEEE Annual Meeting of the Fuzzy Information, USA: IEEE*, 2004: 780-784.
- [13] Yao Yi-yu. An outline of a theory of three-way decision[C]. *Rough Sets and Current Trends in Computing*, Berlin: Springer Berlin Heidelberg, 2012: 1-17.
- [14] Yu Hong, Zhou Qing-feng. A cluster ensemble framework based on three-way decisions[P]. *Proceedings of 8th International Conference on Rough Sets and Knowledge Technology 2013 (RSKT 2013)*, Halifax, Canada, Berlin: Springer Berlin Heidelberg, October 11-14, 2013: 302-312.
- [15] Tang Wei, Zhou Zhi-hua. Bagging-based selective clusterer ensemble[J]. *Journal of Software*, 2005, 16(4): 496-502.
- [16] Yu Hong, Wang Ying. Three-way decisions method for overlapping clustering[P]. *Proceedings of 8th International Conference, Rough Sets and Current Trends in Computing (RSCTC 2012)*, Springer Berlin Heidelberg, 2012: 277-286.

#### 附中文参考文献:

- [8] 周林, 平西建, 徐森, 等. 基于谱聚类的聚类集成算法[J]. *自动化学报*, 2012, 38(8): 1335-1342.
- [15] 唐伟, 周志华. 基于 Bagging 的选择性聚类集成[J]. *软件学报*, 2005, 16(4): 496-502.