

# 控制与决策

Control and Decision

## 一种基于成员选择的簇加权聚类集成算法

徐森, 高婷, 徐秀芳, 许贺洋, 郭乃瑄, 卞学胜, 花小朋, 陈致远

引用本文:

徐森, 高婷, 徐秀芳, 等. 一种基于成员选择的簇加权聚类集成算法[J]. *控制与决策*, 2024, 39(12): 4136-4140.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.1449>

---

## 您可能感兴趣的其他文章

### Articles you may be interested in

#### 基于聚类簇结构特性的自适应综合采样法在入侵检测中的应用

Toward intrusion detection via cluster structure-based adaptive synthetic sampling approach

*控制与决策*. 2021, 36(8): 1920-1928 <https://doi.org/10.13195/j.kzyjc.2019.1672>

#### 基于波段影像统计信息量加权K-means聚类的高光谱影像分类

Algorithm based on band statistical information weighted K-means for hyperspectral image classification

*控制与决策*. 2021, 36(5): 1119-1126 <https://doi.org/10.13195/j.kzyjc.2019.1516>

#### 基于相互邻近度的密度峰值聚类算法

Density peaks clustering based on mutual neighbor degree

*控制与决策*. 2021, 36(3): 543-552 <https://doi.org/10.13195/j.kzyjc.2019.0795>

#### 基于相异性度量选取初始聚类中心改进的K-means聚类算法

Improved K-means clustering algorithm for selecting initial clustering centers based on dissimilarity measure

*控制与决策*. 2021, 36(12): 3083-3090 <https://doi.org/10.13195/j.kzyjc.2020.0554>

#### 一种基于相对密度和决策图的聚类算法

A novel clustering algorithm based on relative density and decision graph

*控制与决策*. 2018, 33(11): 1921-1930 <https://doi.org/10.13195/j.kzyjc.2017.0822>

# 一种基于成员选择的簇加权聚类集成算法

徐森<sup>1†</sup>, 高婷<sup>1</sup>, 徐秀芳<sup>1</sup>, 许贺洋<sup>1</sup>, 郭乃瑄<sup>1,2</sup>, 卞学胜<sup>1</sup>, 花小鹏<sup>1</sup>, 陈致远<sup>1</sup>

(1. 盐城工学院 信息工程学院, 江苏 盐城 224002;

2. 计算机网络和信息集成教育部重点实验室(东南大学), 南京 211189)

**摘要:** 聚类集成算法在数据挖掘和模式识别等领域应用广泛. 现有的聚类集成算法虽取得了显著的进展, 但鲜有同时考虑如何处理冗余成员和关注成员内部多样性的算法. 对此, 设计一种簇的不确定性度量指标, 并提出一种基于成员选择的簇加权聚类集成算法. 首先, 利用平均差异性度量和筛选聚类成员, 并引入信息熵衡量簇的不确定性, 给簇赋予相应的权重; 其次, 在基于成员选择的簇加权共协矩阵和高置信度矩阵的基础上构建增强矩阵; 最后, 在增强矩阵上执行层次聚类算法得到最终的聚类集成结果. 采用多个UCI数据集进行实验, 将所提算法与主流的聚类集成算法进行比较, 实验结果表明, 所提出的算法可以获得更好的聚类集成效果, 且具有较高的鲁棒性和稳定性.

**关键词:** 聚类集成; 数据挖掘; 成员选择; 簇加权; 信息熵; 共协矩阵

中图分类号: TP301.6

文献标志码: A

DOI: 10.13195/j.kzyjc.2023.1449

引用格式: 徐森, 高婷, 徐秀芳, 等. 一种基于成员选择的簇加权聚类集成算法[J]. 控制与决策, 2024, 39(12): 4136-4140.

## A cluster-weighted clustering ensemble algorithm based on member selection

XU Sen<sup>1†</sup>, GAO Ting<sup>1</sup>, XU Xiu-fang<sup>1</sup>, XU He-yang<sup>1</sup>, GUO Nai-xuan<sup>1,2</sup>, BIAN Xue-sheng<sup>1</sup>, HUA Xiao-peng<sup>1</sup>, CHEN Zhi-yuan<sup>1</sup>

(1. Faculty of Information Engineering, Yancheng Institute of Technology, Yancheng 224002, China; 2. Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 211189, China)

**Abstract:** Clustering ensemble algorithms are widely used in fields such as data mining and pattern recognition. Although the existing clustering ensemble algorithms have made significant progress, few algorithms consider how to deal with redundant members and pay attention to the diversity within members at the same time. In this paper, we design an uncertainty metric for clusters, and propose a cluster-weighted clustering ensemble algorithm based on member selection. Firstly, the average difference is used to measure and screen the cluster members, and the uncertainty of the cluster is measured by information entropy, and the corresponding weight is given to the cluster. Then, the enhanced matrix is constructed on the basis of the cluster-weighted co-association matrix and the high-confidence matrix based on member selection. Finally, the hierarchical clustering algorithm is executed on the enhancement matrix to obtain the final clustering ensemble result. Experiments are carried out on multiple UCI datasets, and the proposed algorithm is compared with the mainstream clustering ensemble algorithms, and the experimental results show that the proposed algorithm can obtain better clustering integration effect and has high robustness and stability.

**Keywords:** clustering ensemble; data mining; member selection; cluster weighting; information entropy; co-association matrix

## 0 引言

聚类分析<sup>[1]</sup>是一种无监督学习方法, 在数据挖掘和模式识别等领域应用广泛, 其目标是使同簇内的样

本相似度尽可能高, 异簇间的相似度尽可能低<sup>[2]</sup>. 现有的聚类算法主要包括基于划分、层次、密度、网格、模型等方法<sup>[3-4]</sup>. 由于数据具有多样性特点, 单一的

收稿日期: 2023-10-16; 录用日期: 2024-03-15.

基金项目: 国家自然科学基金项目(62076215, 62301473); 中央高校基本科研业务费专项资金项目(K93-9-2022-03); 江苏省教育厅面上项目(23KJB520039); 江苏省网络与信息安全重点实验室项目(BM2003201); 江苏高校“青蓝工程”项目; 盐城市基础研究计划项目(YCBK2023008); 盐城工学院研究生培养创新工程项目(SJXC23-XY060).

责任编辑: 夏元清.

<sup>†</sup>通讯作者. E-mail: xusen@ycit.cn.

聚类算法常难以充分挖掘其结构. 对此, Strehl等<sup>[5]</sup>提出了聚类集成 (clustering ensemble, CE), 通过组合不同的聚类结果以提高最终的准确性和鲁棒性. 近年来, 许多学者分别针对冗余成员较多和成员内部多样性未能充分挖掘导致聚类结果不佳的问题进行了研究<sup>[6-8]</sup>. Ma等<sup>[9]</sup>提出了选择性聚类集成的解决方案, 根据质量和多样性选择成员参与集成; Rashidi等<sup>[10]</sup>提出了加权证据积累和加权图划分方法评估成员内部可靠性.

虽然上述研究提出了相关改进, 但目前仍缺少可同时考虑处理冗余成员和关注成员内部多样性问题的算法<sup>[11-12]</sup>. 为此, 本文提出一种基于成员选择的簇加权聚类集成算法 (cluster weighted clustering ensemble algorithm based on member selection, CWCEMS). 首先, 通过计算成员的平均差异性度量及筛选成员, 引入信息熵评估簇的不确定性, 并赋予权重; 其次, 构建簇加权共协矩阵和高置信度矩阵<sup>[13]</sup>以获取增强矩阵; 最后, 通过层次聚类算法对增强矩阵进行聚类, 得到最终结果.

本文的主要贡献为: 1) 引入信息熵评估簇的不确定性, 提出簇的不确定性度量指标 (cluster uncertainty measurement metrics, CUMM), 实现关注成员内部多样性的目标. 2) 提出一种基于成员选择的簇加权聚类集成算法, 考虑成员质量的同时关注其内部的多样性, 降低低质量成员对聚类结果的影响.

## 1 成员选择和自增强矩阵

2008年, Fern等<sup>[14]</sup>提出了聚类集成选择 (clustering ensemble selection), 其原理是在聚类集成的过程中增加选择聚类成员步骤, 即从聚类集体  $\Pi$  中精心挑选一个子集, 以形成性能优于  $\Pi$  的小型集体. 这便将传统的聚类集成框架转化成选择性聚类集成框架, 如何评估和选择更好的聚类成员则成为了成员选择的首要问题.

为了实现聚类集成, 现有的集成方法通常会考虑多个成员中两个样本在同一簇中出现的次数构造共协矩阵<sup>[15]</sup>.  $l$  代表样本  $x_a$  和  $x_b$  在成员  $\pi_m$  中是否同簇, 若同簇, 则  $l$  为 1, 否则  $l$  为 0. 文献<sup>[13]</sup>从共协矩阵中提取高置信度信息并将其传播到共协矩阵, 以构建自增强共协矩阵.

## 2 本文算法

### 2.1 成员选择方法

源于信息论的归一化互信息 (normalized mutual information, NMI) 和调整兰德指数 (adjusted rand index, ARI) 是机器学习中常见的评价指标之一, 可

有效衡量成员之间的相似性, 其值越大表示两个聚类结果越相似<sup>[16]</sup>. 本文选取这两个指标衡量成员的相似性和差异性, 通过计算成员的平均差异性<sup>[9]</sup>筛选成员, 如下所示:

$$U_m = \varphi \cdot \text{div}_{1m} + (1 - \varphi) \cdot \text{div}_{2m}. \quad (1)$$

其中:  $U_m$  表示成员  $\pi_m$  的  $U$  值, 度量成员的平均差异性程度;  $\varphi$  表示平衡参数;  $\text{div}_{1m}$  和  $\text{div}_{2m}$  分别表示关于 ARI 和 NMI 的差异值.

### 2.2 簇的不确定性度量指标

信息熵是信息论中的重要概念, 聚类分析中常用其评估簇的不确定性<sup>[8]</sup>. 为了关注簇的多样性, 本文设计一种簇的不确定性度量指标, 如下所示:

$$\begin{cases} H(C_i^m) = - \sum_{j=1}^{n^m} p(C_i^m, C_j^m) \log_2 p(C_i^m, C_j^m), \\ p(C_i^m, C_j^m) = \frac{|C_i^m \cap C_j^m|}{\min(|C_i^m|, |C_j^m|)}, \\ \text{CUMM}(C_i^m) = \frac{\max(H) - H(C_i^m)}{\max(H) - \min(H)}. \end{cases} \quad (2)$$

其中:  $C_i^m$  表示成员  $\pi^m$  中第  $i$  个簇,  $n^m$  表示成员中的簇数量,  $H(C_i^m)$  和  $|C_i^m|$  分别表示簇的信息熵和簇内样本数量,  $|C_i^m \cap C_j^m|$  表示两个簇中具有相同样本的数量,  $H$  表示集体中所有簇的信息熵集合.  $p(C_i^m, C_j^m)$  度量异簇间的相似度, 同时关注它们的大小. CUMM 指标值越高, 说明该簇内部一致性越高.

### 2.3 构建加权的自增强共协矩阵

利用簇的 CUMM 值, 构建簇加权共协矩阵  $B$ , 同时从共协矩阵  $A$  中提取高置信度信息构建高置信度矩阵  $S$ , 以构建增强矩阵  $D$ , 如下所示:

$$\begin{cases} B = \frac{1}{M'} \sum_{m=1}^{M'} l \cdot \text{CUMM}(C_i^m), \\ S = \Psi_{\Omega}(A). \end{cases} \quad (3)$$

$$\begin{cases} \mathcal{L}(D, E, F, Y_1, Y_2) = \\ \text{tr}(D^T \Phi D) + \frac{\beta}{2} \|E\|_F^2 + \langle Y_1, B - D - E \rangle + \\ \frac{\gamma_1}{2} \|B - D - E\|_F^2 + \langle Y_2, D - F \rangle + \frac{\gamma_2}{2} \|D - F\|_F^2; \\ \text{s.t. } \Psi_{\Omega}(E) = 0, F = F^T, 0 \leq F \leq 1. \end{cases} \quad (4)$$

其中:  $M'$  是集成规模;  $\Omega = \{(a, b) | A_{ab} \geq \alpha\}$  表示高置信度元素的位置,  $\alpha$  决定了在多大程度上获取共协矩阵中的条目作为高置信度元素,  $\Psi_{\Omega}(A)$  表示元素映射算子;  $E$  表示误差项,  $F$  表示中间矩阵,  $Y_1$  和  $Y_2$  表示拉格朗日乘子;  $\Phi = Q - S$  表示图拉普拉斯矩阵,  $Q$  表示关于  $S$  的对角矩阵;  $\gamma_1$  和  $\gamma_2$  表示引入增广拉格

朗日项,  $\| \cdot \|_F$  表示矩阵的 Frobenius 范数,  $\beta$  用于平衡误差项. 当  $D$ 、 $E$ 、 $F$ 、 $Y_1$ 、 $Y_2$  中的最大值小于一定阈值时, 此时得到增强矩阵  $D$ . 在增强矩阵上执行平均链接层次聚类算法, 可得到最终的集成结果  $\pi_*$ .

综上所述, 本文算法流程如下所示.

输入: 数据集  $X$ , 簇个数  $k$ , 集成规模  $M'$ ;

输出: 执行平均链接层次聚类算法得到  $\pi_*$ .

1) 运行  $k$ -means 算法  $M$  次, 生成聚类集体  $\Pi = \{\pi_1, \pi_2, \dots, \pi_M\}$ ;

2) 根据式(1)计算各个成员的平均差异性指标, 选出  $M'$  个组成新的聚类集体  $\Pi' = \{\pi'_1, \pi'_2, \dots, \pi'_{M'}\}$ , 并根据式(2)计算簇的 CUMM 值;

3) 分别根据式(3)和(4)构建加权共协矩阵  $B$ 、高置信度矩阵  $S$  和增强矩阵  $D$ .

### 3 实验及结果分析

本节将提出的 CWCEMS 与主流算法进行比较, 选取 ARI、NMI、 $F$  作为评价指标, 并对实验结果进行分析. 实验配置为: Windows 10 64 位操作系统, Intel i5 双核 2.3 GHz 中央处理器, 16 G 内存. 所有实验均在 MatlabR2022a 中运行. 本实验使用来自 UCI 数据库的 7 个数据集作为实验数据集, 分别为 Ecoli、Zoo、Wine、Cancer、Heartstatlog、Vehicle Silhouettes (VS)、

Robotnavigation, 如表 1 所示.

表 1 实验数据表

数据集	样本数	类别数	特征数
Ecoli	336	8	8
Zoo	101	7	16
Wine	178	3	13
Cancer	683	2	9
Heartstatlog	270	2	13
VS	846	4	18
Robotnavigation	5456	4	25

#### 3.1 与其他算法的对比实验

将 CWCEMS 与 EC-CMS<sup>[13]</sup>、LWEA<sup>[8]</sup>、ECPCS-MC<sup>[15]</sup>、ECPCS-HC<sup>[15]</sup>、MDEC-HC<sup>[17]</sup>、MDEC-SC<sup>[17]</sup>、MDEC-BG<sup>[17]</sup> 这 7 种算法进行比较. 为了保证公平公正, 集成规模设置为 20, 分别运行不同的算法各 20 次, 计算它们的平均 F 值、平均 NMI 值和平均 ARI 值, 并引入 “Average” 统计各算法在 7 个数据集上的平均指标值. 结果如表 2 所示: 在 F 指标方面, CWCEMS 在数据集 Ecoli、Zoo、Cancer、Heartstatlog、VS 和 Robotnavigation 上表现最优; 在 NMI 指标方面, CWCEMS 在数据集 Ecoli、Zoo、Wine、Cancer、Heartstatlog 和 Robotnavigation 上表现最优; 在 ARI 指标上, CWCEMS 在数据集 Ecoli、Zoo、

表 2 不同聚类集成算法获得的结果

		Ecoli	Zoo	Wine	Cancer	Heartstatlog	VS	Robotnavigation	Average
CWCEMS	NMI	<b>0.659</b>	<b>0.789</b>	<b>0.423</b>	<b>0.792</b>	<b>0.258</b>	0.137	<b>0.107</b>	<b>0.452</b>
	ARI	<b>0.677</b>	<b>0.805</b>	0.352	<b>0.880</b>	<b>0.340</b>	<b>0.076</b>	<b>0.072</b>	<b>0.457</b>
	F	<b>0.774</b>	<b>0.851</b>	0.596	<b>0.945</b>	<b>0.675</b>	<b>0.393</b>	<b>0.490</b>	<b>0.675</b>
EC-CMS <sup>[13]</sup>	NMI	0.649	0.770	0.411	0.776	0.190	0.135	0.095	0.432
	ARI	0.571	0.681	0.361	0.862	0.252	0.032	0.033	0.399
	F	0.669	0.749	0.586	0.938	0.632	0.360	0.344	0.611
LWEA <sup>[8]</sup>	NMI	0.616	0.769	0.404	0.776	0.186	0.132	0.093	0.425
	ARI	0.455	0.679	0.345	0.862	0.267	0.035	0.036	0.383
	F	0.568	0.747	0.578	0.938	0.631	0.353	0.341	0.594
ECPCS-MC <sup>[15]</sup>	NMI	0.597	0.686	0.391	0.769	0.016	<b>0.164</b>	0.105	0.390
	ARI	0.582	0.615	0.341	0.865	0.026	0.043	0.042	0.359
	F	0.679	0.704	0.580	0.939	0.560	0.391	0.357	0.601
ECPCS-HC <sup>[15]</sup>	NMI	0.576	0.691	0.391	0.776	0.012	0.160	<b>0.107</b>	0.388
	ARI	0.656	0.598	<b>0.378</b>	0.867	0.019	0.052	0.051	0.374
	F	0.763	0.690	<b>0.619</b>	0.939	0.564	<b>0.393</b>	0.384	0.622
MDEC-HC <sup>[17]</sup>	NMI	0.548	0.654	0.291	0.052	0.028	0.129	0.030	0.247
	ARI	0.473	0.464	0.203	-0.041	-0.003	0.012	0.008	0.159
	F	0.587	0.565	0.508	0.654	0.595	0.392	0.486	0.541
MDEC-SC <sup>[17]</sup>	NMI	0.521	0.646	0.230	0.204	0.027	0.183	0.032	0.263
	ARI	0.426	0.446	0.143	0.057	-0.003	0.019	0.016	0.158
	F	0.539	0.550	0.468	0.618	0.592	0.378	0.379	0.504
MDEC-BG <sup>[17]</sup>	NMI	0.531	0.661	0.273	0.102	0.027	0.175	0.036	0.258
	ARI	0.451	0.475	0.187	-0.025	-0.002	0.005	0.004	0.156
	F	0.562	0.573	0.493	0.631	0.590	0.377	0.461	0.527

Cancer、Heartstatlog、VS和Robotnavigation上表现最优,CWCEMS的各项评价指标在“Average”方面均为最优. 总体而言,本文算法在大部分数据集上的结果优于其他算法.

### 3.2 耗时对比实验

面向大规模数据挖掘时,算法的执行效率也是衡量其优劣的重要指标. 考虑到本文算法是在EC-CMS<sup>[13]</sup>的基础上进行改进的,仅与其比较执行效率.

分别运行这两种算法各10次,计算各自的平均运行时间,单位为s,并引入“比率”统计CWCEMS相对于EC-CMS<sup>[13]</sup>的运行时间性能,实验结果如表3所示. 由表3可见,在7个数据集上,算法平均运行时间均满足CWCEMS < EC-CMS,即本文算法的执行速度快于EC-CMS的执行速度;在“比率”方面,本文算法相对于EC-CMS的运行时间均大于零. 总体而言,本文算法的执行效率高于EC-CMS算法.

表3 耗时对比实验结果

	Ecoli	Zoo	Wine	Cancer	Heartstatlog	VS	Robotnavigation
CWCEMS	2.553	0.837	1.368	9.384	2.452	13.102	662.708
EC-CMS <sup>[13]</sup>	7.496	0.955	2.714	71.541	5.861	78.178	789.800
比率	0.341	0.877	0.504	0.131	0.418	0.168	0.839

### 3.3 消融实验

为了更加清晰地观察成员选择和加权的效应,将本文算法过程分为以下3个方面:仅成员选择、仅簇加权、成员选择簇加权. 结果如表4所示,经过成员选择簇加权后,所有数据集的评价指标都有所提高,该策略的各项评价指标在“Average”方面均为最优.

其中在数据集Zoo和Heartstatlog的3个方面的评价指标改善效果最为显著,在数据集Ecoli和VS上稍逊色. 总体可以判断该策略的效果优于仅成员选择策略和仅簇加权策略,稳定性也高于仅成员选择策略和仅簇加权策略,故该策略可在考虑成员质量的同时关注其内部的差异性.

表4 采用不同策略获得的实验结果

		Ecoli	Zoo	Wine	Cancer	Heartstatlog	VS	Robotnavigation	Average
成员选择簇加权	NMI	0.659	<b>0.789</b>	<b>0.423</b>	<b>0.792</b>	<b>0.258</b>	0.137	0.107	<b>0.452</b>
	ARI	0.677	<b>0.805</b>	<b>0.352</b>	<b>0.880</b>	<b>0.340</b>	<b>0.076</b>	<b>0.072</b>	<b>0.457</b>
	F	0.774	<b>0.851</b>	<b>0.596</b>	<b>0.945</b>	<b>0.675</b>	0.393	<b>0.490</b>	<b>0.675</b>
仅成员选择	NMI	<b>0.677</b>	0.781	<b>0.423</b>	<b>0.792</b>	0.244	<b>0.142</b>	0.100	0.451
	ARI	<b>0.705</b>	0.681	<b>0.352</b>	<b>0.880</b>	0.323	0.072	0.050	0.438
	F	<b>0.784</b>	0.749	<b>0.596</b>	<b>0.945</b>	0.666	0.391	0.439	0.653
仅簇加权	NMI	0.526	0.770	0.278	0.562	0.152	0.015	<b>0.117</b>	0.345
	ARI	0.428	0.730	0.231	0.628	0.201	0.001	0.071	0.327
	F	0.634	0.798	0.545	0.844	0.602	<b>0.398</b>	0.481	0.615

### 3.4 不同集成规模下的鲁棒性

为了更加方便地研究集成规模对本文算法的影响,对其做可视化分析. 集成规模为[20, 100],以10递

增,其他设置不变. 实验结果如图1所示,在部分数据集上,本文算法仅有小幅波动,其余均平稳,波动不大;集成规模对本文算法影响不大,鲁棒性较好.

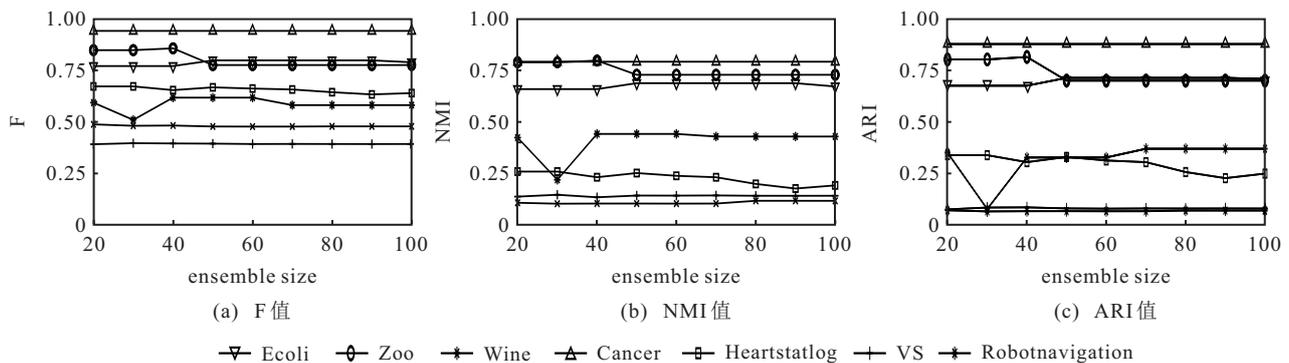


图1 不同集成规模下的评价指标

## 4 结论

本文通过选择高质量成员以及设计簇的不确定性度量指标实现关注冗余成员,并充分挖掘其内部多样性的目标,提出了一种基于成员选择的簇加权聚类集成算法(CWCEMS).在多个数据集上进行的实验结果表明,与主流算法相比,本文算法具有较高的性能和鲁棒性.虽然本文算法表现尚佳,但选择适当的 $U$ 值是进一步需要研究的方向.

### 参考文献(References)

- [1] Chen B W, Xu S, Xu H Y, et al. Structure-aware deep clustering network based on contrastive learning[J]. *Neural Networks*, 2023, 167: 118-128.
- [2] 陈玉明, 蔡国强, 卢俊文, 等. 一种邻域粒 $K$ 均值聚类方法[J]. *控制与决策*, 2023, 38(3): 857-864. (Chen Y M, Cai G Q, Lu J W, et al. A neighborhood granular  $K$ -means clustering method[J]. *Control and Decision*, 2023, 38(3): 857-864.)
- [3] Chen J, Zhu X Z, Liu H W. A mutual neighbor-based clustering method and its medical applications[J]. *Computers in Biology and Medicine*, 2022, 150: 106184.
- [4] 马福民, 宫婷, 杨帆, 等. 基于Zipf分布的网格密度峰值聚类算法[J]. *控制与决策*, 2024, 39(2): 577-587. (Ma F M, Gong T, Yang F, et al. Grid density peak clustering algorithm based on Zipf distribution[J]. *Control and Decision*, 2024, 39(2): 577-587.)
- [5] 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.
- [6] Li D Y, Zhang Z H, Wen G H. Classifier subset selection based on classifier representation and clustering ensemble[J]. *Applied Intelligence*, 2023, 53(18): 20730-20752.
- [7] Wu Y S, Wu R, Liu J F, et al. MetaWCE: Learning to weight for weighted cluster ensemble[J]. *Information Sciences*, 2023, 629: 39-61.
- [8] Huang D, Wang C D, Lai J H. Locally weighted ensemble clustering[J]. *IEEE Transactions on Cybernetics*, 2018, 48(5): 1460-1473.
- [9] Ma T H, Yu T, Wu X G, et al. Multiple clustering and selecting algorithms with combining strategy for selective clustering ensemble[J]. *Soft Computing*, 2020, 24(20): 15129-15141.
- [10] Rashidi F, Nejatian S, Parvin H, et al. Diversity based cluster weighting in cluster ensemble: An information theory approach[J]. *Artificial Intelligence Review*, 2019, 52(2): 1341-1368.
- [11] Zhang D, Yang Y L, Qiu H Q. Two-stage semi-supervised clustering ensemble framework based on constraint weight[J]. *International Journal of Machine Learning and Cybernetics*, 2023, 14(2): 567-586.
- [12] Golalipour K, Akbari E, Motameni H. Cluster ensemble selection based on maximum quality-maximum diversity[J]. *Engineering Applications of Artificial Intelligence*, 2024, 131: 107873.
- [13] Jia Y H, Tao S R, Wang R, et al. Ensemble clustering via co-association matrix self-enhancement[J]. *IEEE Transactions on Neural Networks and Learning Systems*, DOI: 10.1109/TNNLS.2023.3249207.
- [14] Fern X Z, Lin W. Cluster ensemble selection[J]. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 2008, 1(3): 128-141.
- [15] Huang D, Wang C D, Peng H X, et al. Enhanced ensemble clustering via fast propagation of cluster-wise similarities[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, 51(1): 508-520.
- [16] Golalipour K, Akbari E, Hamidi S S, et al. From clustering to clustering ensemble selection: A review[J]. *Engineering Applications of Artificial Intelligence*, 2021, 104: 104388.
- [17] Huang D, Wang C D, Lai J H, et al. Toward multidiversified ensemble clustering of high-dimensional data: From subspaces to metrics and beyond[J]. *IEEE Transactions on Cybernetics*, 2022, 52(11): 12231-12244.

### 作者简介

徐森(1983—),男,教授,博士,硕士生导师,主要研究方向为模式识别与人工智能、机器学习, E-mail: xusen@ycit.cn;

高婷(1997—),女,硕士生,主要研究方向为聚类分析、聚类集成, E-mail: gt714917003@gmail.com;

徐秀芳(1973—),女,高级实验师,硕士,主要研究方向为机器学习、智能信息处理, E-mail: xxf@ycit.cn;

许贺洋(1987—),男,讲师,博士,主要研究方向为机器学习、边缘计算, E-mail: xuheyang124@126.com;

郭乃瑄(1991—),男,讲师,博士,主要研究方向为网络安全、图像处理, E-mail: guonaixuan@ycit.edu.cn;

卞学胜(1991—),男,讲师,博士,主要研究方向为人工智能和医学影像处理, E-mail: xsbian@ycit.edu.cn;

花小朋(1975—),男,副教授,博士,主要研究方向为机器学习、数据挖掘, E-mail: 1462124471@qq.com;

陈致远(2005—),男,学士,主要研究方向为机器学习、智能信息处理, E-mail: 2024151174@qq.com.