

基于图像处理的三支聚类

管文瑞¹, 姜文刚¹, 王平心², 杨习贝³

(江苏科技大学 1.自动化学院; 2.理学院; 3.计算机学院, 江苏 镇江 212100)

摘要:将数字图像处理中模糊锐化算子与三支聚类进行结合,提出了一种基于图像处理的三支聚类算法。该算法通过逆多元二次核函数将数据集的密度量化为灰度值,对数据总体采用模糊与锐化操作,提取锐化后灰度值较高的数据区域,将低密度区域从原始数据中删除。对灰度值较高的数据采用传统的聚类算法得到不同的类簇,然后对每个类簇利用图像模糊算子得到类簇的核心域,锐化算子得到类簇数据边界域,从而获得每个类簇的三支表示。试验采用不同的UCI数据集,通过比较聚类指标 Adjusted Rand Index(ARI), Normalized Mutual Information(NMI)和 Adjusted Mutual Information(AMI),验证了该聚类算法的有效性。

关键词:三支聚类;非参数核密度估计;数字图像处理;模糊锐化

中图分类号:TP391 **文章编号:**1005-9830(2023)05-0643-08

DOI:10.14177/j.cnki.32-1397n.2023.47.05.010

Three-way clustering based on digital image processing

Guan Wenrui¹, Jiang Wengang¹, Wang Pingxin², Yang Xibei³

(1.School of Automation; 2.School of Sciences; 3.School of Computer,
Jiangsu University of Science and Technology, Zhenjiang 212003, China)

Abstract: A three-way clustering algorithm based on image processing is proposed by combining blurring and sharpening operations in digital image processing with three-way clustering. The proposed algorithm quantifies the density of the dataset into gray values through inverse multivariate quadratic kernel function. By blurring and sharpening operations for the university, we delete the samples with low density and obtain samples with high density. Then, different clusters are produced by traditional clustering algorithm for samples with high density. For each cluster, the core region and fringe region are obtained through blurring operation and sharpening operation, respectively. The experiments on different UCI datasets verify that the clustering algorithm is stable and effective by comparing the clustering indexes adjusted rand index(ARI), normalized mutual information(NMI) and adjusted mutual information(AMI).

Key words: three-way clustering; nonparametric kernel density estimation; digital image processing; blurring and sharpening

收稿日期:2022-12-29 修回日期:2023-07-11

基金项目:国家自然科学基金(62076111; 61773012);江苏省研究生科研与实践创新计划项目(KYCX23_3883)

作者简介:管文瑞(2000-),男,硕士生,主要研究方向:智能控制, E-mail:gwr0615@163.com;通讯作者:王平心(1980-),男,博士,副教授,主要研究方向:粗糙集、粒计算, E-mail:pingxin_wang@hotmail.com。

引文格式:管文瑞,姜文刚,王平心,等.基于图像处理的三支聚类[J].南京理工大学学报,2023,47(5):643-650.

投稿网址:http://zrxuebao.njust.edu.cn

聚类是研究样本分类问题的一种统计分析方法,同时也是数据挖掘的一个重要算法。聚类的目的是将数据集划分成多个类簇,使得在同一个类簇中的对象尽可能的相似,不同类簇中的对象不相似。传统聚类算法用一个集合表示一个类,不同的类簇之间有清晰的边界,但在实际划分中常常会遇到信息不充分的情况,如果将数据对象强行划分到某一类簇,会造成误分类的概率增加,导致聚类精度降低。针对传统聚类方法的不足,基于三支决策理论^[1],Yu 提出了三支聚类的概念^[2],其利用三元思想将每个类簇表示为核心域、边缘域和琐碎域^[3],使得聚类结果的描述更为合理。一般来说构造三个区域需要阈值,而实际样本中数据的不确定性使得固定阈值划分的区域效果较差。对此,Jia 等^[4]提出了一种自适应的阈值确定方法,通过核心域与边缘域范围得到 Pvi 指标,将其最优解作为阈值。一些学者将其优化为无阈值划分方法,如 Yu 等^[5]基于改进 DBSCAN 的方法得到三支聚类结果。Wang 等^[6]采用形态学中腐蚀膨胀的思想,将收缩后的区域定为核心域,扩张得到边缘域,克服了固定阈值的局限性。

数字图像处理,是将二维图像表示成数组或矩阵形式,以像素为基本计算单位进行运算。其中将聚类分析的方法在数字图像处理中有着广泛的应用,如基于聚类分析算法的图像分割研究^[7];Ali 等^[8]将数字图像处理的方法反向应用于聚类分析,构建了 BS3WC 聚类分析算法。本文引入数字图像处理^[9]中低通高通滤波器的思想,提出基于数字图像处理的三支聚类方法。使用核密度估计将数据集转化为灰度值,将低通滤波(模糊化)得到稳定的高密度区域定义为核心域,将高通滤波(锐化)得到稳定的低密度区域定义为琐碎域,中间部分则为边缘域。

1 预备知识

1.1 三支聚类

设数据集 $U = \{x_1, x_2, \dots, x_n\}$ 包含 n 个样本对象,传统的聚类方法是使用一个集合表示一个类,即寻找一组集合 C_1, C_2, \dots, C_k 满足

$$\begin{cases} C_i \neq \phi, i = 1, \dots, k \\ \bigcup_{i=1}^k C_i = U \\ C_i \cap C_j = \phi, i \neq j \end{cases}$$

式中: k 为类簇的个数。上述条件要求一个对象要么属于某个类簇,要么不属于某个类簇,聚类的结果具有清晰的边界,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。

为了克服传统聚类方法存在的问题,Yu 将三支决策^[1]的思想引入到聚类中,提出了三支聚类方法^[2]。与硬聚类不同的是,三支聚类将数据点与类簇之间的关系分为:确定属于、可能属于以及不属于三类。基于这一思想,三支聚类用 $Co(C_i)$ 、 $Fr(C_i)$ 、 $Tr(C_i)$ 表示一个类,即核心域、边界域和琐碎域,其中 $Co(C)$ 中的样本点一定属于类 C , $Fr(C)$ 中的样本可能属于类 C , $Tr(C)$ 中的对象一定不属于类 C 。由于琐碎域可通过公式 $Tr(C_i) = U - Co(C_i) - Fr(C_i)$ 进行表示,三支聚类的结果可以表示为

$$C = \{ (Co(C_1), Fr(C_1)), (Co(C_2), Fr(C_2)) \dots, (Co(C_k), Fr(C_k)) \}$$

三支聚类用核心域与边界域来表示一个类,这种聚类方法充分考虑了那些由于信息不充分而无法确定类簇归属的样本对象,降低了决策风险,被广泛应用于属性约简^[10]、不确定性分析^[11]等领域。

1.2 核密度估计

核密度估计(Kernel density estimation)是根据数据样本的分布情况来获得其总体特征,即为样本的概率密度函数估计,根据概率密度核函数,便可以得到数据分布性质,如数据的聚集区域。常见的密度估计由式(1)给出

$$\rho_i = \sum_{x_j \in V} K \left(\frac{\|x_j - x_i\|}{h} \right) \quad (1)$$

式中: K 为非负的单调递减核函数,变量 x_i 和 x_j 为样本点和目标点。 V 是一组样本。 h 是用于控制规模的带宽参数,带宽越大所形成的密度曲线越平滑,但会掩盖峰值较小的密度特征。对于核函数 K ,文献[12]分析说明核函数对密度曲线形态的影响不大。因此带宽是影响曲线的重要参数,一般通过交叉验证法,经验法确定带宽^[13]。聚类分析也是通过分析有限的样本数据来得到数据点的特征,将核密度估计作为聚类分析的前处理方法^[14],具有良好的可扩展性,对于离群数据或噪声数据不敏感,并且可以对任意形状的数据进行挖掘^[15,16]。

1.3 图像处理中的模糊与锐化

模糊(Blurring)运算即低通空间滤波器,在图像处理中用于削减噪声的干扰,并保证图像信息的完整性,又称为图像的平滑^[9](Image smoothing)。一般使用的模糊方法有两类:空域法和频域法。空域法是运用模块化运算,提取目标像素点邻域的像素值特性,根据不同的模糊算子进行滤波。频域法是运用正交变换的操作,利用噪声对应的高频信息的特点进行滤波^[17]。本文使用空域法进行模糊化操作,空域法又称空间域平滑滤波,常用的方法有均值滤波、高斯滤波和中值滤波等。设输入图像为 U_{ol} ,模糊化图像为 U_{bl} ,模糊运算的定义可表示为

$$Lpf(x_i) = \frac{1}{N} \sum_{(x_m) \in S} f(x_m) \quad (2)$$

式中: S 为 x_i 邻域像素点的集合, N 为 S 中像素的个数。

对图像 U_{ol} 所有像素点进行模糊得到模糊图像 U_{bl} 。如图 1(a) 的模糊化操作,高密度数据边缘灰度值会进行衰减,而核心部分不受影响,由此应用于三支聚类,便可以缩减出稳定的核心域 $Co(C_i)$ 。而边缘区域以及低密度区域灰度值也会进行适当降低,使得图像边缘的涵盖范围适当扩大。

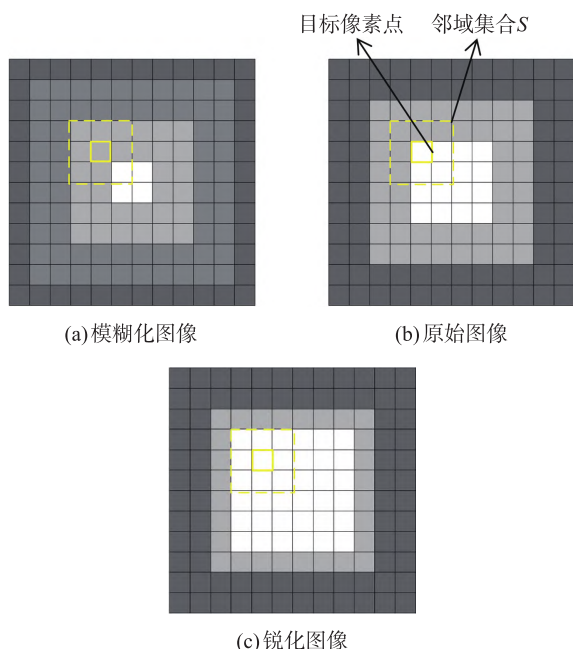


图 1 图像模糊锐化处理示例

锐化(Sharpening)运算即高通空间滤波器,在图像中通过定义不同的微分算子,且该算子的响应与灰度的变化幅度相关^[9]。因此锐化操作

会将灰度图中的边缘区域灰度值提高,而对于衰减信号或低频信号强度降低。本文中锐化操作中添加了模糊化的图像特征^[17],设锐化后的图像为 U_{sp} ,定义锐化运算为

$$Spf(x_i) = f(x_i) + k(f(x_i) - Lpf(x_i)) \quad (3)$$

式中: k 为锐化效果参数, k 值越大,低灰度值与高灰度值的分离程度越大,使得低密度区域的数据点更为疏远。对原始图像 U_{ol} 所有像素点进行锐化得到锐化图像 U_{sp} 。将该操作映像到三支聚类当中,锐化(高通)滤波器强化了数据集的高密度的边缘区域,并且弱化了数据密度较低的区域,对于离群点(低频信号)进行削减。如图 1(c) 的锐化操作,高密度与低密度交界处填补为高密度区域,且数据边界不会扩张,因此锐化操作也与数据集三支聚类的期望相符合。

2 基于图像处理的三支聚类算法

2.1 基于核密度估计的二值化图像

根据输入数据集的特征进行采样与量化。采样为决定数字化图像的分辨率,即待处理的聚类分析图像的像素数量。量化是根据核密度估计的方法将各个像素点赋予灰度值。由采样与量化便可以将数据集转换为灰度图像。核密度估计是应用较为广泛的密度测量方法,其定义如公式(1)所示。对于核函数 K ,本文采用如下的逆多元二次核函数(Inverse multiquadric kernel)^[18],表达式如下

$$K(x, x_i) = \frac{1}{\sqrt{\|x - x_i\|^2 + c^2}} \quad (4)$$

式(4)是一种局部性强的函数,形状为钟形曲线。采用该函数具有两大优势,第一个是模糊系数 c 为像素框的对角线长度,可以对于不同的数据特征自适应地改变参数,没有需要用户指定的系数,提高了其可用性。第二个是引用其中模糊系数 c ,避免了低密度区域中的像素(数据之间距离过大)赋予灰度值过低,而高密度区域中的像素赋予的灰度值过高。对于模糊系数 c 的计算由式(5)给出

$$c = \sqrt{x_{1w}^2 + x_{2w}^2 + \dots + x_{nw}^2} \quad (5)$$

根据逆多元二次核函数可以得到数据点的相对密度。当数据点密度较低时,则分配该点对应的像素点较小的灰度值(一种接近黑色的颜色),当

数据点密度较高时,则分配该点对应的像素点较大的灰度值(一种接近白色的颜色)。选取某点附近 n 个数据点分别计算核函数得到与其对应的密度值,将密度值进行求和并量化至 $[0, 1]$ 区间内。选取的数据点 n 一般为数据集规模的十分之一。将核函数应用于样本数据集,二维样本数据集特征是均值为 100, 方差为 5 的正态分布,可得到灰度值图像如图 2 所示,可见灰度值图像与三支聚类的语义较为贴切,结构上与三支聚类一一对应,白色像素区域与核心域关联,而外部浅灰色区域与边界域相关联。在下一节中,应用图像处理中的模糊锐化,对图像进行进一步处理。

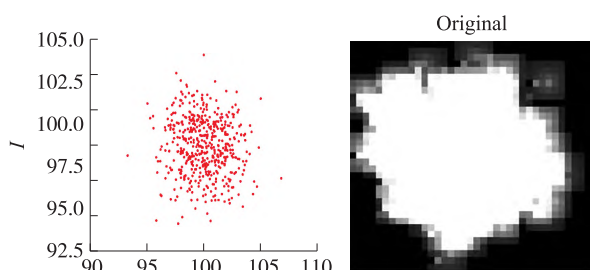


图 2 量化二维数据点图像

2.2 数据图像模糊锐化处理

这一步将对原始的灰度图像进行模糊锐化处理。设低通滤波器(Low pass filter)为 $Lpf(x_i)$, 集合 $S(f_1, \dots, f_m)$ 为像素点 $f(x_i)$ 的邻域像素点, 利用式(2)可得原始图像 U_{ol} 模糊化操作结果。

模糊操作的主要思想是将像素点 $f(x_i)$ 相邻 r 半径内^[19] 像素点的均值灰度返回至像素点 $f(x_i)$, r 设置越大, 灰度图像模糊化效果就越为平滑。本文中试验结果的领域半径 r 定为 5 倍的参数 c , 即 $r = 5c$ 。如果像素点 $f(x_i)$ 位于高密度边缘, 且上方灰度值小于 1, 进行模糊化使得其灰度值减小, 脱离了高密度区域。由此得知进行均值模糊化操作后, 数据集中高密度区域进行了缩减, 与文献[6]中腐蚀膨胀的方法具有相似之处。

锐化操作将原始图像与模糊图像的差值补充在原始图像中以达到锐化效果, 其定义如式(3)所示。锐化操作效果较为自然, 避免了锐化边界灰度值突出等问题。图 3 给出了图像的模糊和锐化效果图, 其中(b)为原始图像, (a)与(c)分别为模糊和锐化效果图。

2.3 提取数据特征

锐化也为高通滤波器, 可以将其作为聚类分析的前处理方法。锐化操作可以提取锐化后灰度值较高的数据区域, 该区域为有效聚类区域, 对该

区域运行聚类可以有效降低灰度值较低区域对聚类结果的影响。用一个例子来说明这一问题。图 4(a)为原始数据图, 采用传统聚类方法获得聚类结果图 4(c)所示。图 4(b)为对原始数据图锐化提取的有效区域, 其聚类结果如图 4(d)所示。通过对比图 4(c)与 4(d), 发现提取锐化特征后使得聚类对象确, 聚类结果有效性提高。

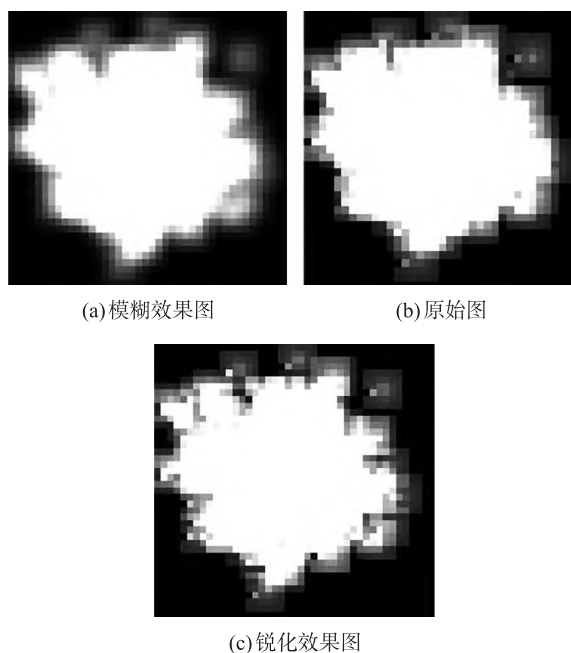


图 3 模糊锐化示例图

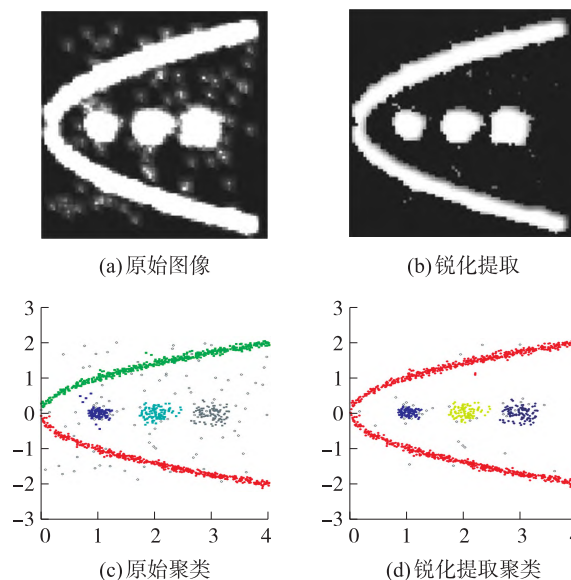


图 4 锐化提取示例图

对锐化提取的有效区域, 采用传统聚类算法得到原始的类簇 $C_k = \{C_1, C_2, \dots, C_k\}$ 。为了较好地刻画数据的不确定性, 利用模糊与锐化得到每个类

的核心域与边界域。由于将原始图像 U_o 经过模糊化,高密度区域得到缩减,剩下稳定的核心区域,定义该区域为类簇的核心域 $Co(C_i)$,表示为

$$Co(c_i) = \{x_i \in U | x_i \in C_i \wedge Lpf(x_i) = 1\} \quad (6)$$

模糊操作所提取的特征,说明了该区域模糊前后,像素值始终为 1,该情况为目标像素点邻域 $S(f_1, \dots, f_m)$ 全为 1,即为稳定的高密度区域。而锐化操作后,图像中高密度区域得到扩展,这意味着扩张部分与核心域关系较为紧密,定义该区域为边界域 $Fr(C_i)$,表示为

$$Fr(c_i) = \{x_i \in U | x_i \in C_i \wedge Spf(x_i) = 1 \wedge Lpf(x_i) \neq 1\} \quad (7)$$

该区域经过锐化填补像素值后,扩张为高密度区域,而剩下的部分始终为低密度区域,说明其像素值过小,与核心域联系过于疏远,定义为琐碎域 $Tr(C_i)$,表示为

$$Tr(c_i) = \{x_i \in U | x_i \in C_i \wedge Spf(x_i) \neq 1 \wedge Lpf(x_i) \neq 1\} \quad (8)$$

通过上述操作,获得了每个类簇的核心域、边界域和琐碎域,其中边界域中数据与该类簇关系紧密,而琐碎域包含了与类簇 C_i 关系很弱数据,噪声以及离群点等数据。因此琐碎域并不是该类簇的一部分,由此获得将数据集 $U = \{x_1, x_2, \dots, x_n\}$ 的三支聚类结果

$$C = \{(Co(C_1), Fr(C_1)), (Co(C_2), Fr(C_2)) \dots, (Co(C_k), Fr(C_k))\} \quad (9)$$

2.4 基于图像处理的三支聚类(3WDIP)算法流程

基于图像处理的三支聚类算法基本思想是采用数字图像处理中模糊与锐化的思想提取数据集的特征,即对原始数据进行核密度估计,使得二维数据可以生成可视化灰度图像,高维数据仅得到数据点处灰度值。首先对数据总体采用模糊与锐化操作,得到总体数据特征,即稳定的高密度区域以及稳定的低密度区域。提取锐化后灰度值较高的数据区域,将低密度区域从原始数据中删除。对灰度值较高的数据采用传统的聚类算法得到不同的类簇,这样可以有效降低灰度值较低区域对聚类结果的影响。为了较好地刻画数据的不确定性,本文利用对每个类簇进行模糊与锐化得到每个类的核心域与边界域,由此获得将数据集的三支聚类结果。算法 1 给出了基于图像处理的三支聚类(3WDIP)算法过程。

2267048036002

算法 1 3WDIP 算法

输入:数据集 U ,簇数目 k ,领域半径 r 。

输出: $C = \{(Co(C_1), Fr(C_1)), (Co(C_2), Fr(C_2)) \dots, (Co(C_k), Fr(C_k))\}$

步骤 1 根据数据集规模采样,并得到核函数中参数 c_o 。

步骤 2 使用核密度估计进行量化得到原始灰度数据 U_o 。

步骤 3 对原始灰度 $f(x, y)$ 进行模糊操作 $Lpf(x_i) = \frac{1}{N_{(x_m) \in S}} \sum f(x_m)$ 得到模糊灰度 U_{bl} 。

步骤 4 根据原始灰度数据 U_o 和模糊灰度 U_{bl} 得到锐化灰度 U_{sp} 。

步骤 5 对锐化提取区域数据进行聚类得到原始聚类结果 C_i (本文分别采用 k-means 与 DPC)。

步骤 6 提取 $Co(c_i) = \{x_i \in U | x_i \in C_i \wedge Lpf_{(m,n)} = 1\}$,得到 C_i 的核心域。

步骤 7 提取 $Fr(c_i) = \{x_i \in U | x_i \in C_i \wedge Spf_{(m,n)} = 1 \wedge Lpf_{(m,n)} \neq 1\}$,得到 C_i 的边界域。

步骤 8 得到最终聚类结果: $C = \{(Co(C_1), Fr(C_1)), (Co(C_2), Fr(C_2)) \dots, (Co(C_k), Fr(C_k))\}$

3 试验分析

为了对本文算法的有效性进行验证,本文选取了 12 组 UCI 数据集进行试验。数据集信息如表 1 所示,其中部分数据添加了噪声数据。

表 1 试验所用数据集 UCI

Data sets	Instance	Features	Classes
Iris	150	4	3
Libras	360	90	15
Seeds	210	7	3
Segmentation	2310	19	6
Wine	178	13	3
Forest	523	27	4
Glass	214	9	6
Breast	106	9	6
Card	2126	21	10
Page	5473	10	5
Thyroid	215	5	3
Pathbased	300	2	3

首先对原始数据进行核密度估计,得到数据点处灰度值,然后对数据集进行模糊和锐化操作。提取锐化后的高密度区域,采用 k-means 获得硬聚类结果,然后对每个类簇进行模糊与锐化得到

每个类的核心域与边界域,由此获得将数据集的三支聚类结果并计算聚类指标调整兰德系数 ARI,标准化互信息 NMI,调整互信息 AMI。ARI 的范围为 $[-1,1]$,值越大意味着聚类结果与真实

情况越吻合,NMI 与 AMI 指标值域为 $[0,1]$,值越大两个聚类结果的相近程度越接近。试验结果如表 2 所示。为了对比本文算法的效果,表 2 还给出了 UCI 数据集用 k-means 聚类的结果。

表 2 3WDIP 嵌入 k-means 试验结果

Data sets	3WDIP-kmeans			k-means		
	ARI	AMI	NMI	ARI	AMI	NMI
Iris	0.770±0.017	0.818±0.032	0.823±0.031	0.719±0.052	0.744±0.051	0.747±0.051
Libras	0.417±0.006	0.602±0.013	0.655±0.010	0.285±0.039	0.510±0.041	0.568±0.034
Seeds	0.762±0.026	0.781±0.007	0.785±0.007	0.711±0.014	0.696±0.011	0.699±0.011
SSegmentation	0.465±0.051	0.644±0.053	0.646±0.053	0.374±0.031	0.532±0.038	0.534±0.039
Wine	0.415±0.003	0.555±0.003	0.568±0.001	0.366±0.005	0.420±0.002	0.426±0.002
Forest	0.493±0.090	0.593±0.054	0.597±0.053	0.383±0.113	0.467±0.081	0.470±0.080
Glass	0.241±0.032	0.417±0.067	0.449±0.065	0.157±0.023	0.266±0.025	0.297±0.024
Breast	0.317±0.017	0.504±0.026	0.556±0.024	0.322±0.083	0.465±0.018	0.507±0.019
Card	0.238±0.028	0.416±0.025	0.422±0.025	0.162±0.019	0.319±0.025	0.325±0.025
Page	0.255±0.031	0.478±0.027	0.479±0.027	0.054±0.011	0.100±0.015	0.101±0.015
Thyroid	0.420±0.023	0.573±0.042	0.581±0.083	0.579±0.065	0.488±0.035	0.488±0.035
Pathbased	0.664±0.038	0.744±0.054	0.747±0.034	0.464±0.024	0.545±0.022	0.548±0.020

将对应数据的均值相差,得到指标差值对比图 5。图中负值表示使用 3WDIP 后聚类效果下降。

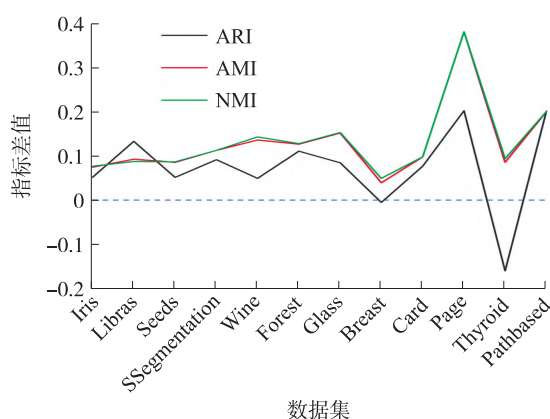


图 5 聚类指标差值图

为了进一步验证算法的有效性,对上述数据集锐化提取的高密度区域,使用 DPC 聚类获得硬聚类结果,然后对每个类簇进行模糊与锐化得到每个类的核心域与边界域。图 6 为使用 3WDIP 嵌入 DPC 后得到的聚类结果与直接采用 DPC 算法得到的聚类结果在 ARI、NMI 和 AMI 上的对比图。

观察表 2 和图 6 的对比试验数据,可得出以下结论:

(1)通过对比 NMI 标准互信息发现,3WDIP 算法应用于 K-means 时,各项数值增幅较为稳定。其中数据集 Page 在运用了 3WDIP 算法后,ARI 和 NMI 的值为 0.2856 和 0.4939,而仅运行 K-means 的两个数值为 0.060 1 和 0.116 6。这种改进归因于锐化提取中消除了噪音以及离群点的干扰,仅对核心域和边界域进行聚类,明确了各个类簇之间的边界。

(2)将其应用于 DPC 算法时,各项数值平均增幅比 K-means 要高,这归因于 3WDIP 算法中模糊与锐化处理。灰度值与 DPC 中密度值意义相关联,在经过模糊锐化,各个类簇的边界域和核心域的灰度值得到补充,而琐碎域的灰度值得到下沉,使得类簇与外部之间关联性降低,提高了类内的聚集程度。而数据集 Breast 未能得到改进,是由于类簇内部数据分布不均,数据内部存在鞍点,使得 3WDIP 降低了鞍点左右的灰度值,导致其应用于 DPC 时效果不佳,而在 K-means 中稳定发挥。

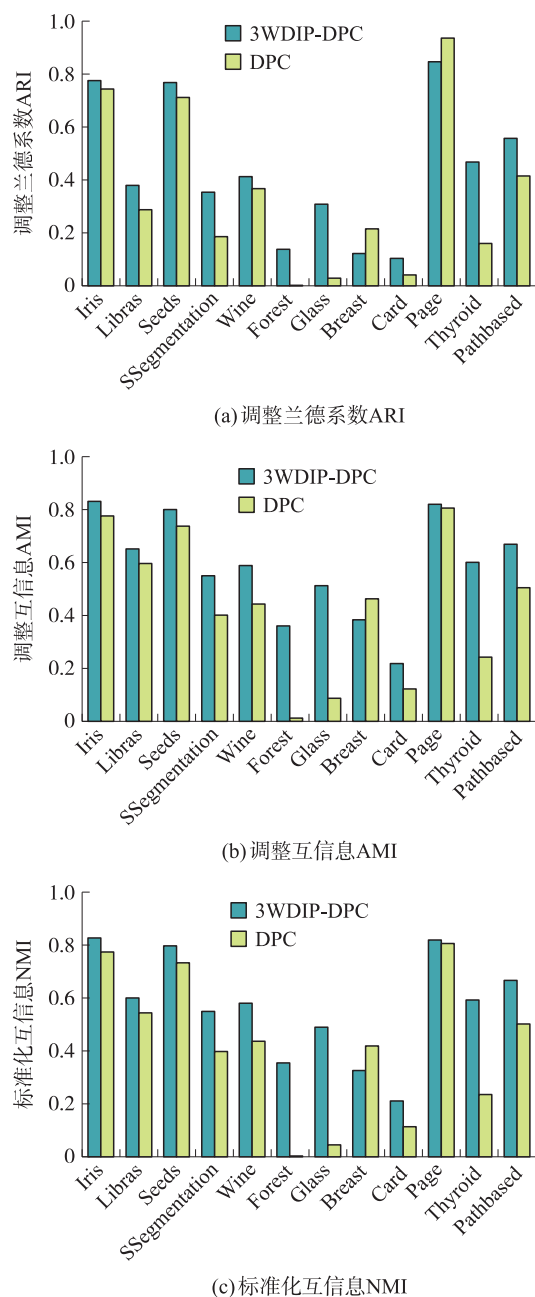


图 6 3WDIP 嵌入 DPC 试验结果

综上所述,3WDIP 算法能较好地对数据特征进行处理,能应用于不同的聚类方式,并且能较为稳定的提升聚类性能,其三支聚类结果可解决现实中不确定性的聚类问题。

4 结论

本文将数字图像处理中模糊与锐化的思想与三支聚类进行结合,给出了一种基于图像处理的三支聚类算法。这算法通过提取锐化后灰度值较高的数据区域,对灰度值较高的数据采用传统的

聚类算法得到不同的类簇,然后对每个类簇利用图像模糊算子得到类簇的核心域,锐化算子得到类簇数据边界域,从而获得每个类簇的三支表示。试验结果表明,本文的算法将核心域和边界域突出显示,提高了各个类簇之间的分离程度,使得各项聚类性能稳定的增加。这算法仍存在需要改进的地方,如应对类簇内部数据分散不均而产生的鞍点,3WDIP 算法会对这区域进行削弱,这种错判使得类内分散性提高。此外,在模糊化操作中,虽然引用了自适应的可达半径进行模糊操作,还需丰富评价模糊化所需数据参数,如何进一步提取数据特征进行优化也是下一步的研究工作之一。

参考文献:

- [1] Yao Yiyu. Three-way decisions with probabilistic rough sets [J]. Information Sciences, 2010, 180 (3): 341-353.
- [2] Yu Hong. A framework of three-way cluster analysis [C]//International Joint Conference on Rough Sets. Cham, Switzerland: Springer, 2017: 300-312.
- [3] Yu Hong, Zhang Cong, Wang Guoyin. A tree-based incremental overlapping clustering method using the three-way decision theory [J]. Knowledge-Based Systems, 2016, 91: 189-203.
- [4] Jia Xiuyi, Rao Ya, Li Weiwei, et al. An automatic three-way clustering method based on sample similarity [J]. International Journal of Machine Learning and Cybernetics, 2021, 12(5): 1545-1556.
- [5] Yu Hui, Chen Luyuan, Yao Jingtiao, et al. A three-way clustering method based on an improved DBSCAN algorithm [J]. Physica A: Statistical Mechanics and Its Applications, 2019, 535: 122289.
- [6] Wang Pingxin, Yao Yiyu. CE3: A three-way clustering method based on mathematical morphology [J]. Knowledge-Based Systems, 2018, 155: 54-65.
- [7] 唐利明, 田学全, 黄大荣, 等. 结合 FCMS 与变分水平集的图像分割模型 [J]. 自动化学报, 2014, 40(6): 1233-1248.
- [8] Tang Liming, Tian Xuequan, Huang Darong, et al. Image segmentation model combined with FCMS and variational level set [J]. Acta Automatica Sinica, 2014, 40(6): 1233-1248.
- [8] Ali B, Azam N, Yao Jingtiao. A three-way clustering approach using image enhancement operations [J].

- International Journal of Approximate Reasoning, 2022, 149:1-38.
- [9] Gonzalez R C, Woods R E. Digital Image Processing [M]. 2nd ed. Boston: Wesley Longman Publishing Co, Inc, 1992.
- [10] 张昭琴, 徐泰华, 鞠恒荣, 等. 基于粒度的加速求解约简策略[J]. 南京理工大学学报, 2021, 45(4): 401-408.
- Zhang Zhaoqin, Xu Taihua, Ju Hengrong, et al. Accelerative solving reduct strategy based on granularity [J]. Journal of Nanjing University of Science and Technology, 2021, 45(4): 401-408.
- [11] 张倚萌, 贾修一, 唐振民. 基于条件信息熵的区间集决策信息表不确定性度量[J]. 南京理工大学学报, 2019, 43(4): 393-401.
- Zhang Yimeng, Jia Xiuyi, Tang Zhenmin. Uncertainty measurement for interval set decision information tables based on conditional information entropy [J]. Journal of Nanjing University of Science and Technology, 2019, 43(4): 393-401.
- [12] Harpole J K. How bandwidth selection algorithms impact exploratory data analysis using kernel density estimation [D]. Lawrence, USA: University of Kansas, 2013.
- [13] Sheikhpour R, Sarram M A, Sheikhpour R. Particle swarm optimization for bandwidth determination and feature selection of kernel density estimation based classifiers in diagnosis of breast cancer [J]. Applied Soft Computing, 2016, 40: 113-131.
- [14] 李存华, 孙志挥, 陈耿, 等. 核密度估计及其在聚类算法构造中的应用[J]. 计算机研究与发展, 2004, 41(10): 1712-1719.
- Li Cunhua, Sun Zhihui, Chen Geng, et al. Kernel density estimation and its application to clustering algorithm construction [J]. Journal of Computer Research and Development, 2004, 41(10): 1712-1719.
- [15] 淦文燕, 李德毅. 基于核密度估计的层次聚类算法[J]. 系统仿真学报, 2004, 16(2): 302-305.
- Gan Wenyan, Li Deyi. Hierarchical clustering based on kernel density estimation [J]. Acta Simulata Systematica Sinica, 2004, 16(2): 302-305.
- [16] 郭婧, 耿海军, 吴勇. 基于菌群优化的 K 均值聚类算法研究[J]. 南京理工大学学报, 2021, 45(3): 314-319.
- Guo Jing, Geng Haijun, Wu Yong. K-means clustering algorithm based on bacteria foraging optimization [J]. Journal of Nanjing University of Science and Technology, 2021, 45(3): 314-319.
- [17] Shah A, Azam N, Ali B, et al. A three-way clustering approach for novelty detection [J]. Information Sciences, 2021, 569: 650-668.
- [18] Breiman L, Meisel W, Purcell E. Variable kernel estimates of multivariate densities [J]. Technometrics, 1977, 19(2): 135-144.
- [19] 叶丽珠, 郑冬花, 隋栋, 等. 面向高维数据的人工蜂群优化密度峰值聚类[J]. 南京理工大学学报, 2022, 46(3): 291-297.
- Ye Lizhu, Zheng Donghua, Sui Dong, et al. Artificial bee colony optimized density peak clustering for high-dimensional data [J]. Journal of Nanjing University of Science and Technology, 2022, 46(3): 291-297.