



基于可拓距的改进k-means聚类算法

赵燕伟, 朱芬, 桂方志, 任设东, 谢智伟, 徐晨

引用本文:

赵燕伟, 朱芬, 桂方志, 等. 基于可拓距的改进k-means聚类算法[J]. 智能系统学报, 2020, 15(2): 344–351.

ZHAO Yanwei, ZHU Fen, GUI Fangzhi, et al. Improved k-means algorithm based on extension distance[J]. *CAAI Transactions on Intelligent Systems*, 2020, 15(2): 344–351.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.201811020>

您可能感兴趣的其他文章

加权PageRank改进地标表示的自编码谱聚类算法

An autoencoder spectral clustering algorithm for improving landmark representation by weighted PageRank
智能系统学报. 2020, 15(2): 302–309 <https://dx.doi.org/10.11992/tis.201904021>

结合谱聚类的标记分布学习

Label distribution learning based on spectral clustering
智能系统学报. 2019, 14(5): 966–973 <https://dx.doi.org/10.11992/tis.201809019>

负载敏感的云任务三支聚类评分调度研究

Load-aware score scheduling of three-way clustering for cloud task
智能系统学报. 2019, 14(2): 316–322 <https://dx.doi.org/10.11992/tis.201710004>

基于最大最小距离的高光谱遥感图像波段选择

Maximin distance algorithm-based band selection for hyperspectral imagery
智能系统学报. 2018, 13(1): 131–137 <https://dx.doi.org/10.11992/tis.201703023>

应用k-means算法实现标记分布学习

Label distribution learning based on k-means algorithm
智能系统学报. 2017, 12(3): 325–332 <https://dx.doi.org/10.11992/tis.201704024>

适合大规模数据集的增量式模糊聚类算法

Incremental fuzzy (c+p)-means clustering for large data
智能系统学报. 2016, 11(2): 188–199 <https://dx.doi.org/10.11992/tis.201507013>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.201811020

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20190829.1004.004.html>

基于可拓距的改进 k-means 聚类算法

赵燕伟¹, 朱芬¹, 桂方志¹, 任设东², 谢智伟¹, 徐晨¹

(1. 浙江工业大学特种装备制造与先进加工技术教育部/浙江省重点实验室, 浙江 杭州 310014; 2. 浙江工业大学计算机科学与技术学院, 浙江 杭州 310014)

摘要: 针对现有聚类算法在初始聚类中心优化过程中存在首个初始聚类中心点落于边界非密集区域的不足, 导致出现算法聚类效果不均衡问题, 提出一种基于可拓距优选初始聚类中心的改进 k-means 算法。将样本经典距离向可拓区间映射, 并通过可拓侧距计算方法得到可拓左侧距及可拓右侧距; 引入平均可拓侧距概念, 将平均可拓左侧距和平均可拓右侧距分别作为样本密集度和聚类中心疏远度的量化指标; 在此基础上, 给出初始聚类中心选取准则。通过与传统 k-means 聚类算法进行对比, 结果表明改进后的 k-means 聚类算法选取的初始聚类中心分布更加均匀, 聚类效果更好, 尤其在对高维数据聚类时具有更高的聚类准确率和更好的均衡性。

关键词: 可拓距; k-means 聚类算法; 缩放因子; 初始聚类中心; 密集度; 疏远度

中图分类号: TP181 **文献标志码:** A **文章编号:** 1673-4785(2020)02-0344-08

中文引用格式: 赵燕伟, 朱芬, 桂方志, 等. 基于可拓距的改进 k-means 聚类算法 [J]. 智能系统学报, 2020, 15(2): 344-351.

英文引用格式: ZHAO Yanwei, ZHU Fen, GUI Fangzhi, et al. Improved k-means algorithm based on extension distance[J]. CAAI transactions on intelligent systems, 2020, 15(2): 344-351.

Improved k-means algorithm based on extension distance

ZHAO Yanwei¹, ZHU Fen¹, GUI Fangzhi¹, REN Shedong², XIE Zhiwei¹, XU Chen¹

(1. Key Lab of Special Purpose Equipment and Advanced Manufacturing Technology, Ministry of Education & Zhejiang Province, Zhejiang University of Technology, Hangzhou 310014, China; 2. College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310014, China)

Abstract: An improved k-means algorithm optimizing the initial cluster centers based on extension distance was proposed to solve several problems that lead to clustering imbalance of the algorithm, such as the poor quality of initial cluster center selection or the first initial cluster center easily falling into the non-dense area of the data boundary. First, the classical distance of the sample was mapped onto the extension interval, and the extension left-side and right-side distances were obtained using the extension distance calculation method. Then, the average extension side distance was determined, and the extension left-side and right-side distances were taken as the quantitative indicators of sample density and cluster center distance, respectively. Subsequently, the selection criteria of the initial cluster center were given. Finally, compared with the traditional k-means algorithm, the improved k-means algorithm obtained higher clustering accuracy and better balance, particularly in high-dimensional data clustering.

Keywords: extension distance; k-means clustering algorithm; scaling factor; initial cluster center; intensity; alienation

聚类是数据分析的重要手段, 将数据集分为若干类, 使得簇内紧密且相似性大, 簇与簇之间

有明显区别, 使得相似性最小, 在数据挖掘、图像处理等领域被广泛应用^[1-4]。k-means 聚类算法是一种常用的动态聚类算法, 具有聚类速度快, 操作简单, 效率高等特点, 但其同时存在对初始聚类中心点较敏感、全局搜索能力弱的缺点, 使得

收稿日期: 2018-11-26. 网络出版日期: 2019-08-29.

基金项目: 国家自然科学基金项目 (51875524); 浙江省公益技术应用研究计划项目 (2017C31072).

通信作者: 赵燕伟 (1959-). E-mail: ywz@zjut.edu.cn.

聚类效率低,准确性差。因此,很多学者为得到稳定、高效的聚类效果,对 k-means 聚类算法的局限性进行了改进与研究。

何熊熊等^[5]提出一种基于密度和网格的簇心可确定聚类算法,通过网格对象的密度值进行划分以完成聚类;李亚等^[6]引入轮廓系数作为聚类效果评价指标,并将改进的算法应用到台折损率的计算中;邢长征等^[7]针对聚类结果对孤立点敏感的问题,提出了一种基于平均密度优化初始聚类中心的 k-means 算法;张天骐等^[8]提出一种基于 k-means 聚类改进的软扩频信号伪码序列盲估计算法,利用相似度从分段数据中寻找初始聚类中心,并通过平均轮廓系数估计规模数;李晓瑜等^[9]提出了一种基于 Hadoop 的分布式改进 k-means 算法,通过引入 Canopy 算法初始化 k-means 算法的聚类中心;Tzortzis G 等^[10]提出了 MinMax k-means 算法,在初始化不好的情况下通过赋予方差权重以优化目标提高聚类效果;Laith Mohammad Abualigah 等^[11]将 k-means 算法采用和声搜索方法进行优化并应用于文本聚类中,提高了聚类精度;Li Yanyan 等^[12]提出一种基于粒子群优化的 k-means 算法,并将其应用在岩体不连续数据分类中;Khanmohammadi 等^[13]为克服对初始聚类中心的敏感问题,提出了一种混合 k-谐波均值和重叠 k 均值算法的混合方法来克服缺点。以上改进算法都起到了较好的聚类效果,但在初始聚类中心选取问题上仍存在首个初始聚类中心点落于稀疏边界的缺陷。

基于此,本文提出一种基于可拓距的改进 k-means 聚类算法,将可拓学思想与 k-means 算法有效的结合,通过引入可拓侧距和缩放因子,对首个初始聚类中心点进行优化,选出一组最优初始聚类中心点,并通过仿真对比检验本文改进算法的可行性。通过实验验证,该方法具有更好的聚类效果。

1 基本知识

1.1 可拓距相关知识

可拓学是以蔡文教授为首的我国学者们创立的新学科,近年来,可拓学在计算机、人工智能、检测、控制等领域进行的应用取得了良好的成绩^[14]。其中,可拓距在实例检索领域应用较为广泛,通过可拓距构造关联函数,依据样本间关联度识别案例类别^[15-19],显著提高了案例检索效率与正确率。

在经典数学中,当点在区间内时,默认点与区

间的距离为零,即“类内即为同”,无法将事物内部的“质变”、“量变”表达出来。因此,为了描述类内事物(区间内的点)的区别,在可拓学中规定:实轴上任意一点 x 与区间 $X_0=\langle a, b \rangle$ 之距为^[14]

$$\rho(x, X_0) = |x - \frac{a+b}{2}| - \frac{b-a}{2} = \begin{cases} a-x, & x \leq \frac{a+b}{2} \\ x-b, & x \geq \frac{a+b}{2} \end{cases} \quad (1)$$

当实例点在区间中点时,该实例最符合要求; $\langle a, b \rangle$ 可以是开区间、闭区间、半开区间。

但在实际工作中,却不全如此,例如,误差要求越小越好;成本要求越低越好;性价比越高越好;洗衣机的洗净率越大越好等,一般而言,实例点并不是越在区间中点越好,因此在可拓距的基础上,引入可拓侧距的概念^[14]:

给定区间 $X_0=\langle a, b \rangle$, $x_0 \in (a, \frac{a+b}{2})$, 称

$$\rho_l(x, x_0, X_0) = \begin{cases} a-x, & x \leq a \\ \frac{b-x_0}{x-x_0}(x-a), & x \in \langle a, x_0 \rangle \\ a-x_0, & x \geq x_0 \\ x-b, & x \geq x_0 \end{cases} \quad (2)$$

为 x 与区间 X_0 关于 x_0 的左侧距。

给定区间 $X_0=\langle a, b \rangle$, $x_0 \in (\frac{a+b}{2}, b)$, 称

$$\rho_r(x, x_0, X_0) = \begin{cases} a-x, & x \leq x_0 \\ \frac{a-x_0}{b-x_0}(b-x), & x \in \langle x_0, b \rangle \\ b-x_0, & x \geq b \\ x-b, & x \geq b \end{cases} \quad (3)$$

为 x 与区间 X_0 关于 x_0 的右侧距。

1.2 k-means 聚类算法基本原理

k-means 聚类算法基本思想是将样本划分成多个类,使得各簇内对象具有尽可能大的相似度,同时使簇间的相似度尽可能的小^[20]。k-means 聚类算法的处理过程如下:

- 1) 从数据集 X 中随机选择 k 个对象,分别作为 k 个类别的初始聚类中心;
- 2) 计算剩余每个对象与各个聚类中心的距离,并将其划分到距离最近的子类中;
- 3) 重新计算每个子类中所有对象的平均值,将其作为新的聚类中心。

重复上述过程,直到聚类中心不再改变^[21]。

1.3 k-means 聚类算法的不足分析

k-means 算法中,对于 k 个初始中心点的选取是随机完成的,而初始中心点选取的不同会导致不同的聚类效果,从而引起聚类结果的不稳定性。针对该不足,一些学者提出用密集度、差异度等^[22-23]概念对初始聚类中心进行优化,都无法避免初始聚类中心点落在边界非密集区域,因此本文将从初始中心点选取方面对 k-means 算法提出相应的改进措施。

2 可拓距改进的 k-means 聚类算法

2.1 基本思想

为便于表述,首先定义距离区间、距离可拓侧距、距离平均可拓侧距3个概念。对于 n 个样本的集合 $X = \{x_1, x_2, \dots, x_n\}$,其中 x_i 为 m 维向量($i = 1, 2, \dots, n$),有如下定义:

定义1 样本集合 X 的距离区间 Z 为

$$Z = [A, B] = [\min(D), \max(D)] \quad (4)$$

其中, $D = \left\{ d \mid d = \sqrt{\sum_{p=1}^m (x_i^p - x_j^p)^2} \right\}$ 为两两样本间距离集合。

定义2 根据式(2)和式(3)定义两样本 x_i 和 x_j ($i \neq j$) 的距离 d 对区间 Z 的左、右侧距分别为

$$\text{左侧距: } \rho_l^{(i,j)} = \rho_l(d, A, Z) = \begin{cases} A-d, & d < A \\ A_z, & d = A \\ d-B, & d > A \end{cases}$$

$$\text{其中: } A_z = \rho_l(A, A, Z) = \begin{cases} 0, & A \notin Z \\ A-B, & A \in Z \\ 0 \otimes (A-B), & A \notin Z \text{ 且 } A \in Z. \end{cases} \quad (5)$$

$$\text{右侧距: } \rho_r^{(i,j)} = \rho_r(d, B, Z) = \begin{cases} A-d, & d < B \\ B_z, & d = B \\ d-B, & d > B. \end{cases}$$

$$\text{其中: } B_z = \rho_r(B, B, Z) = \begin{cases} 0, & B \notin Z \\ A-B, & B \in Z \\ 0 \otimes (A-B), & B \notin Z \text{ 且 } B \in Z. \end{cases} \quad (6)$$

定义3 根据定义2可计算所有两两样本的可拓侧距,则平均左、右侧距为

$$\bar{\rho} = \frac{\sum_{j=i+1}^n \sum_{i=1}^{n-1} \rho^{(i,j)}}{C_n^2} \quad (7)$$

其中, C_n^2 表示从 n 个样本中任意取2个样本的组合数。

针对传统 k-means 算法初始中心点随机选取所引起聚类算法稳定性差问题,现有的改进算法虽取得了一定的效果,但仍无法避免初始聚类中心点落在边界非密集区域如图1所示,取样本间距离最小值所对应中心坐标作为首个初始聚类点,因下一初始聚类中心点的选取决定于首个初始中心的位置,当该点位于边界非密集区域,既降低了剩余初始聚类中心点质量,又会出现最终聚类集合中样本数为0或1的情况,使得聚类效果不均衡。

初始聚类中心的选取,不仅要求分布在较密集的范围,还需要保证各初始聚类中心尽可能分散。针对上述问题,利用可拓距中数据分布情

况对左右侧距平均值的影响,将经典平均距离映射为两个平均侧距值,如图2所示,其中平均左侧距值相对中心点靠左分布,平均右侧距值相对中心点靠右分布。特别指出,当数据在区间对称分布时,左右平均可拓侧距值重合于一点。将可拓平均左侧距 $\bar{\rho}_l$ 作为衡量密集度指标,可拓平均右侧距 $\bar{\rho}_r$ 作为衡量疏远度指标,首个大于可拓平均左侧距 $\bar{\rho}_l$ 所对应中心坐标作为第一个聚类中心,下一待选取初始聚类中心点需满足与各已确定初始聚类中心点间可拓距均大于可拓右侧平均距 $\bar{\rho}_r$,方可作为聚类中心点。

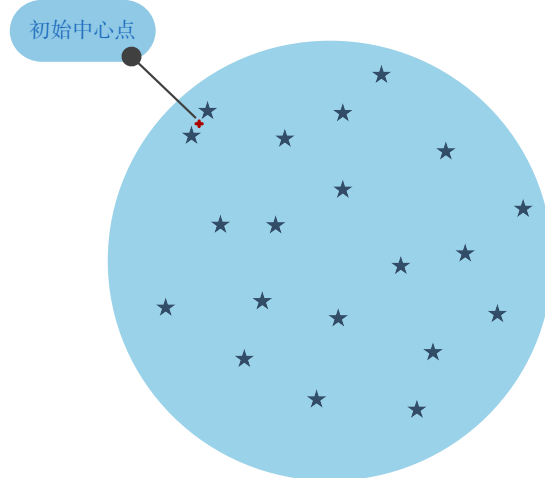


图1 初始中心点展示图

Fig. 1 The display map of first initial center point

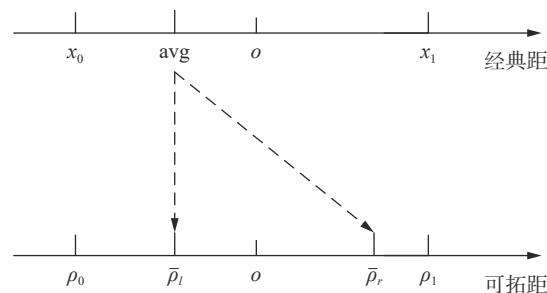


图2 经典距向可拓距映射

Fig. 2 Mapping of classical distances to extension distances

当选出初始聚类中心点数未达到所要求个数时,引入缩放因子 η 如式(7)所示,对可拓平均右侧距进行缩放,选出满足个数的初始聚类中心点。

$$\eta = \begin{cases} 1 + \frac{C_n^2 - k'}{C_n^2} & k' \neq K \\ 1, & k' = K \end{cases} \quad (8)$$

其中, k' 为每次遍历后,所获得的初始聚类中心个数; K 为指定聚类中心数。

最后按传统聚类算法进行聚类。选取的一组最优初始聚类中心点克服了中心点出现于边界非密集区域缺陷,最大限度分布在密集区且各聚类中心点均匀分布。

2.2 改进 k-means 算法初始聚类中心选取流程

根据上述思想,得到改进 k-means 算法初始聚类中心选取的具体实施步骤如下:

1)按式(4)计算出两两样本间距离及等效密集距离区间 $[A,B]$;

2)按式(5)和式(6)将距离映射为可拓左侧距 $\rho_l^{(i,j)}$ 及可拓右侧距 $\rho_r^{(i,j)}$,将 $\rho_l^{(i,j)}$ 按从小到大顺序依次排序,同时按式(7)计算样本间可拓平均左侧距 $\bar{\rho}_l$ 及可拓平均右侧距 $\bar{\rho}_r$;

3)遍历排序好的可拓距,将其中首个大于样本间可拓平均左侧距 $\bar{\rho}_l$ 的可拓距对应中心点坐标作为第一个初始聚类中心。

4)计算排好序可拓距中下一个值对应中心点坐标并依次计算出其与已确定的初始聚类中心的可拓距,将其与样本平均可拓右侧距 $\bar{\rho}_r$ 进行比较,若其均大于 $\bar{\rho}_r$,则该中心点坐标作为下一个初始聚类中心;否则重新执行步骤4)。

5)如果遍历一次后,初始聚类中心未达到 K ,则按式(8)计算出缩小因子 η ,动态缩小样本平均可拓右侧距 $\bar{\rho}_r$,重新回到步骤3)。

6)若聚类中心数达到 K 时,则完成初始聚类中心的选取。

3 实验与分析

为了验证本文所提出算法的有效性,将本文算法与传统 k-means 算法及文献[22-23]所提出的改进聚类算法进行对比分析。

实验所用的测试数据集为 UCI 数据库中用于测试聚类的 Iris 数据集和 Wine 数据集,各数据的特征如表1所示。

表1 各数据集的基本特征
Table 1 Characteristics of datasets

数据集	样本个数	样本维数	分类数
Iris	150	4	3
Wine	179	13	3

基于本文提出算法对 Iris、Wine 数据进行初始中心点选取,特别指出,为了观察初始聚类中心点选取位置的大体远近及分散程度,本文为节省篇幅,只展示数据集两属性的二维图3和图4。从图3和图4中可看出本文提出改进算法选取的初始聚类中心点,对低维与高维数据选取的初始聚类中心点,相较于其他改进算法分布更均匀,位于边界区域初始聚类中心点,其周围数据点也相对密集。

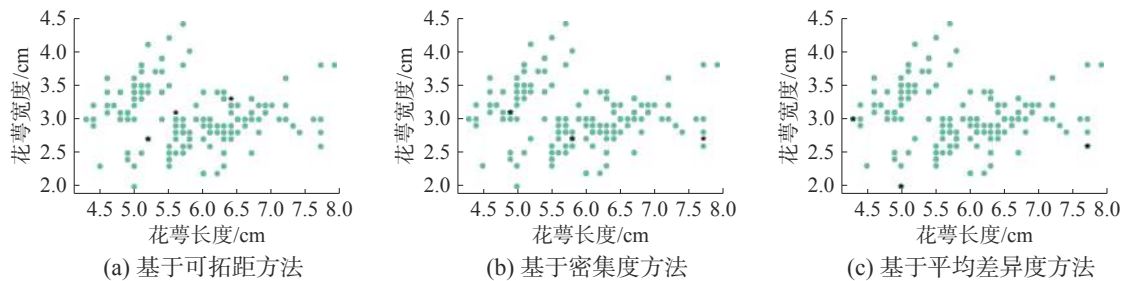


图3 基于 Iris 数据的初始聚类中心点分布对比

Fig. 3 Comparison of distribution of initial cluster centers based on Iris dataset

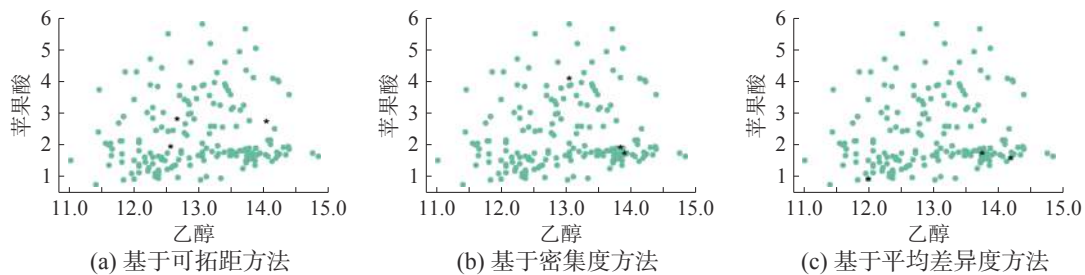


图4 基于 Wine 数据的初始聚类中心点分布对比

Fig. 4 Comparison of distribution of initial cluster centers based on Wine dataset

为了定量描述初始聚类中心点选取的质量,本文先将所选中心点聚类,将其与样本实际聚类

情况和现有改进算法聚类情况进行对比,其聚类效果图5和图6。

其中,图5与图6中3种颜色代表3个聚类簇,可看出其他两种改进算法,由于初始中心点分布不均匀且有位于稀疏边界的中心点,从而导致对两组样本集聚类的结果存在一定误差,但本文提出

的改进算法,虽然仍存在误差,但误差的大小有所缩小,尤其在对更高维的Wine数据集聚类时,优势更明显。

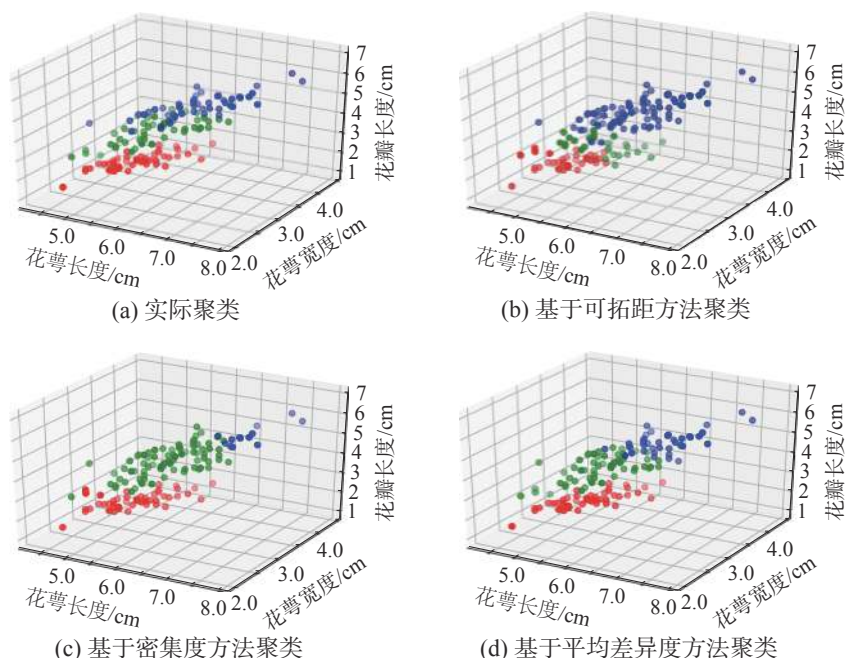


图5 基于Iris数据聚类对比图

Fig. 5 Comparison of clustering based on Iris dataset

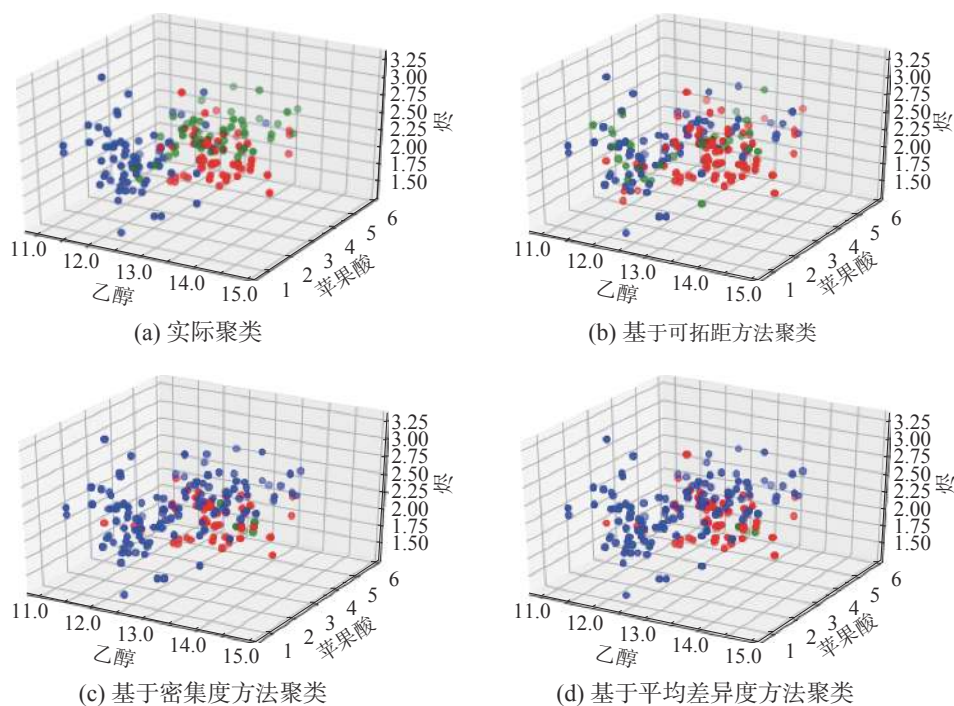


图6 基于Wine数据聚类对比

Fig. 6 Comparison of clustering based on Wine dataset

为了定量衡量算法的有效性,本文采用分类正确率指标,即被正确分到对应类别的样本个数

与总样本个数比值以及聚类均衡性指标,即各类中样本个数与实际相应类中样本个数的方差,进

行衡量。

对初始聚类中心点聚类后分类正确率统计如表 2 所示。对初始聚类中心点聚类后各簇类样本的个数统计如表 3 所示。由表 3 得出改进前后算法的均衡性指标如表 4 所示。从表 2 与表 4 中可看出当数据集样本维数不高时, 本文算法与改进

算法在聚类的均衡性上差异较小, 但当维数增高时, 现有改进算法聚类均衡性和分类准确性会降低, 导致个别类中样本较多, 相反也会出现有些类中样本过少的情况, 使得之后聚类中心点更新时, 较少样本数据的中心点在迭代更新时坐标变化不明显, 从而导致最终聚类效果差。

表 2 改进前后算法聚类正确率指标

数据集	文献[17]改进算法	文献[18]改进算法	传统k-means算法	本文提出算法
Iris	87.36	85.24	82.37	86.79
Wine	61.05	63.07	60.12	65.28

表 3 改进前后算法聚类后各簇类样本个数

Table 3 The number of samples in each cluster

数据	实际各簇数	本文提出算法各簇数	文献[17]改进算法各簇数	文献[18]改进算法各簇数
Iris	[50, 50, 50]	[40, 29, 81]	[53, 84, 13]	[61, 39, 50]
Wine	[59, 71, 48]	[27, 78, 73]	[118, 53, 7]	[116, 4, 58]

表 4 改进前后算法聚类均衡性指标

Table 4 Balance of clustering algorithm before and after

均衡性指标	实际聚类	本文提出算法聚类	文献[17]改进算法聚类	文献[18]改进算法聚类
Iris	0	22.4	29.1	8.98
Wine	0	23.8	42.77	51.11

4 结束语

本文提出一种基于可拓距的改进 k-means 聚类算法, 将样本间平均距离映射为可拓距上对应两个平均侧距, 作为初始聚类中心选取的约束范围, 克服了首个选取的初始中心点可能位于较疏散边界的缺陷。实验表明在高维数据集的情况下, 基于可拓距的改进 k-means 算法选取的初始聚类中心点聚类具有较高的正确率和均衡率, 表明选取的初始聚类中心点较优, 更适合大数据背景下的海量、高维数据聚类。

本文对 k-means 聚类算法初始聚类中心的选取展开了研究, 由于时间和个人能力有限, 还存在不足之处有待后续研究中深化、提高: 1) 本文算法对非数值型的数据集样本如何衡量相似度, 并没有给出解决方法。2) 文中提到的缩放因子并不能自适应的变化, 对算法的性能有所影响。3) 如何引入关联函数代替距离计算相似度, 是值得深入研究的问题。4) 在理论上和实用上的意义及价值。5) 进一步深入研究本课题的建议。

参考文献:

- [1] 于佐军, 秦欢. 基于改进蜂群算法的 k-means 算法 [J]. 控制与决策, 2018, 33(1): 181–185.
YU Zuojun, QIN Huan. k-means algorithm based on improved artificial bee colony algorithm[J]. Control and decision, 2018, 33(1): 181–185.
- [2] HE Hong, TAN Yonghong. Automatic pattern recognition of ECG signals using entropy-based adaptive dimensionality reduction and clustering[J]. *Applied soft computing*, 2017, 55: 238–252.
- [3] PENG Chong, KANG Zhao, XU Fei, et al. Image projection ridge regression for subspace clustering[J]. *IEEE signal processing letters*, 2017, 24(7): 991–995.
- [4] RÖTHLISBERGER V, ZISCHG A P, KEILER M, et al. Identifying spatial clusters of flood exposure to support decision making in risk management[J]. *Science of the total environment*, 2017, 598: 593–603.
- [5] 何熊熊, 管俊轶, 叶宣佐, 等. 一种基于密度和网格的簇心可确定聚类算法 [J]. 控制与决策, 2017, 32(5): 913–919.
HE Xiongiong, GUAN Junyi, YE Xuazuo, et al. A dens-

- ity-based and grid-based cluster centers determination clustering algorithm[J]. *Control and decision*, 2017, 32(5): 913–919.
- [6] 李亚, 刘丽平, 李柏青, 等. 基于改进 K-Means 聚类 and BP 神经网络的台区线损率计算方法 [J]. *中国电机工程学报*, 2016, 36(17): 4543–4551.
- LI Ya, LIU Liping, LI Baiqing, et al. Calculation of line loss rate in transformer district based on improved K-Means clustering algorithm and BP neural network[J]. *Proceedings of the CSEE*, 2016, 36(17): 4543–4551.
- [7] 邢长征, 谷浩. 基于平均密度优化初始聚类中心的 k-means 算法 [J]. *计算机工程与应用*, 2014, 50(20): 135–138.
- XING Changzheng, GU Hao. K-means algorithm based on average density optimizing initial cluster centre[J]. *Computer engineering and applications*, 2014, 50(20): 135–138.
- [8] 张天骐, 杨强, 宋玉龙, 等. 一种 K-means 改进算法的软扩频信号伪码序列盲估计 [J]. *电子与信息学报*, 2018, 40(1): 226–234.
- ZHANG Tianqi, YANG Qiang, SONG Yulong, et al. Blind estimation PN sequence in soft spread spectrum signal of improved K-means algorithm[J]. *Journal of electronics & information technology*, 2018, 40(1): 226–234.
- [9] 李晓瑜, 俞丽颖, 雷航, 等. 一种 K-means 改进算法的并行化实现与应用 [J]. *电子科技大学学报*, 2017, 46(1): 61–68.
- LI Xiaoyu, YU Liying, LEI Hang, et al. The parallel implementation and application of an improved K-means algorithm[J]. *Journal of University of Electronic Science and Technology of China*, 2017, 46(1): 61–68.
- [10] TZORTZIS G, LIKAS A. The MinMax k-means clustering algorithm[J]. *Pattern recognition*, 2014, 47(7): 2505–2516.
- [11] ABUALIGAH L M, KHADER A T, AI-BETAR M A. Unsupervised feature selection technique based on harmony search algorithm for improving the text clustering[C]//*Proceedings of the 7th International Conference on Computer Science and Information Technology*. Amman, Jordan, 2016: 1–6.
- [12] LI Yanyan, WANG Qing, CHEN Jianping, et al. K-means algorithm based on particle swarm optimization for the identification of rock discontinuity sets[J]. *Rock mechanics and rock engineering*, 2015, 48(1): 375–385.
- [13] KHANMOHAMMADI S, ADIBEIG N, SHANEHBANDY S. An improved overlapping k-means clustering method for medical applications[J]. *Expert systems with applications*, 2017, 67: 12–18.
- [14] 杨春燕, 蔡文. 可拓学 [M]. 北京: 科学出版社, 2014.
- [15] 管凤旭. 基于流形学习及可拓分类器的手指静脉识别研究 [D]. 哈尔滨: 哈尔滨工程大学, 2010.
- GUAN Fengxu. Research on finger vein recognition based on manifold learning and extension classifier[D]. Harbin: Harbin Engineering University, 2010.
- [16] 赵燕伟, 苏楠, 张峰, 等. 基于可拓实例推理的产品族配置设计方法 [J]. *机械工程学报*, 2010, 46(15): 146–154.
- ZHAO Yanwei, SU Nan, ZHANG Feng, et al. Configuration design method for product family based on extension case reasoning[J]. *Journal of mechanical engineering*, 2010, 46(15): 146–154.
- [17] 叶永伟, 张帆, 王运. 基于可拓距的起重机产品配置方法设计 [J]. *中国制造业信息化*, 2012, 41(23): 24–27.
- YE Yongwei, ZHANG Fan, WANG Yun. The crane products configuration design based on extension distance[J]. *Manufacturing information engineering of China*, 2012, 41(23): 24–27.
- [18] NOUAOURIA N, BOUKADOUM M. Case retrieval with combined adaptability and similarity criteria: application to case retrieval nets[C]//*Proceedings of the 18th International Conference on Case-Based Reasoning*. Research and Development. Alessandria, Italy, 2010: 242–256.
- [19] 赵燕伟, 任设东, 陈尉刚, 等. 基于改进 BP 神经网络的可拓分类器构建 [J]. *计算机集成制造系统*, 2015, 21(10): 2807–2815.
- ZHAO Yanwei, REN Shedong, CHEN Weigang, et al. Extension classifier construction based on improved BP neural network[J]. *Computer integrated manufacturing systems*, 2015, 21(10): 2807–2815.
- [20] 李敏. K-means 算法的改进及其在文本聚类中的应用研究 [D]. 无锡: 江南大学, 2018.
- LI Min. The research and application of text clustering based on improved K-means algorithm[D]. Wuxi: Jiangnan University, 2018.
- [21] 杨明极, 马池, 王姬, 等. 一种改进 K-means 聚类的 FCMM 算法 [J]. *计算机应用研究*, 2019, 36(7): 2007–2010.
- YANG Mingji, MA Chi, WANG Ya, et al. Algorithm named FCMM to improve K-means clustering algorithm[J]. *Application research of computers*, 2019, 36(7): 2007–2010.
- [22] 韩俊, 谈健, 黄河, 等. 基于改进 K-means 聚类算法的供电块划分方法 [J]. *电力自动化设备*, 2015, 35(6): 123–129.
- HAN Jun, TAN Jian, HUANG He, et al. Power-supplying block partition based on improved K-means clustering algorithm[J]. *Electric power automation equipment*, 2015, 35(6): 123–129.
- [23] 李武, 赵娇燕, 严太山. 基于平均差异度优选初始聚类中心的改进 K-均值聚类算法 [J]. *控制与决策*, 2017,

32(4): 759–762.

LI Wu, ZHAO Jiaoyan, YAN Taishan. Improved K-means clustering algorithm optimizing initial clustering centers based on average difference degree[J]. Control and decision, 2017, 32(4): 759–762.

作者简介:



赵燕伟, 教授, 博士生导师, 博士, 主要研究方向为可拓设计理论与方法、物流系统智能配送与优化调度、数字化产品现代设计。出版教材 4 部, 多次获得国家自然科学基金项目资助等。发表学术论文 100 余篇。



朱芬, 硕士研究生, 主要研究方向为可拓设计。



桂方志, 博士研究生, 主要研究方向为可拓设计。

全国第 17 届可拓学年会

“全国第 17 届可拓学年会”将于 2020 年 8 月 14—16 日在江苏南京召开。会议由中国人工智能学会主办, 由中国人工智能学会可拓学专业委员会、南京审计大学、《智能系统学报》编辑部、广东工业大学可拓学与创新方法研究所等单位联合承办。会议将交流近年来可拓学的理论和应用研究成果, 讨论可拓学发展方向、学科建设和人才培养等问题。为此诚邀国内外相关领域的专家学者、管理人员和工程技术人员参加会议。

本次会议征文范围为可拓学理论、方法与应用方面的最新研究成果, 包括但不限于:

- 1) 可拓论: 基元理论、可拓集、可拓逻辑。
- 2) 可拓创新方法: 可拓模型建立方法, 拓展分析方法, 可拓变换方法, 优度评价方法, 矛盾问题求解方法等。
- 3) 可拓工程: 可拓智能、可拓设计、可拓控制、可拓检测、可拓决策、可拓管理, 以及可拓学与其他学科(如计算机、数学、自动化、审计、经济、教育、教学等)的交叉研究成果。

征文要求:

1) 文笔精练、资料可靠, 字数不超过 7000 字, 并附中英文摘要和关键词。

2) 论文应是新的研究成果, 严格按照科技论文规范撰写, 且未在学术会议或刊物上发表。

会议征文中被大会录用的论文, 需要作者参加会议交流, 再由可拓学学术委员会评审出优秀论文, 推荐到《智能系统学报》、《数学的实践与认识》等核心期刊和《广东工业大学学报》(中国科技核心期刊)“可拓论坛”正式发表。

欢迎相关领域的研究人员、高校师生、企业家、工程技术人员、以及一切爱好和有志于可拓学研究的朋友踊跃投稿。

本届大会的征文采用 E-mail 投稿的方式, 投稿邮箱: extenics@vip.163.com

重要日期:

征文截止日期: 2020 年 5 月 30 日

录用通知日期: 2020 年 6 月 30 日

会议时间: 2020 年 8 月 14—16 日

联系方式:

联系地址: 广东工业大学可拓学与创新方法研究所

联系人: 李剑明 汤龙

E-mail: extenics@vip.163.com

联系电话: 020-39322973, 传真: 020-39322019

会议网址: <http://extenics.gdut.edu.cn/>