文章编号:1007-5321(2020)05-0021-06 DOI:10.13190/j.jbupt.2020-017

半监督聚类目标下粒子群算法的分析与改进

孙 艺[1], 夏启钊[2]
(1.北京邮电大学 计算机学院(国家示范性软件学院), 北京 100876; 2.北京邮电大学 国际学院, 北京 100876)

摘要: 传统粒子群算法的优点较为明显,但是随着环境复杂度的增高,传统算法的聚类中心敏感度升高,空聚类过多,类标号对聚类结果的影响不足等问题日趋严重. 为此,提出了一种改进算法,以半监督 K 均值聚类为目标,以自适应 K 值的方式,随机地计算初始化聚类中心,并根据均值聚类算法的需要编码成粒子,同时引入软性约束概念重新构造目标函数;最后使用改进后的算法进行寻优. 所提出的粒子群算法改进了自适应参数,引入了免疫扰动和混沌扰动 2 种扰动方式,同时应用了退火策略和动态聚类策略. 实验结果表明,该算法在很大程度上解决了上述问题.

关 键 词: 半监督; K 均值; 信息熵; 扰动; 退火策略
中图分类号: TP301.6 **文献标志码**: A

Analysis and Improvement of Semi-Supervised *K*-means Clustering Based on Particle Swarm Optimization Algorithm

SUN Yi[1], XIA Qi-zhao[2]
(1. School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing 100876, China; 2. International School, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Traditional particle swarm optimization has obvious advantages, but with increased complexity of the environment. When the traditional algorithm is used, the sensitivity of the clustering center is increased, there are too many empty clusters, and the performance of the class label has insufficient influence on the clustering results. An improved algorithm is proposed, which aims at semi-supervised *K*-means clustering; first, the clustering center is initialized by random calculation in an adaptive *K*-value method, and the particles are encoded according to the needs of the mean clustering algorithm. At the same time, the objective function is reconstructed with the concept of soft constraints, and finally the improved algorithm is used for optimization. The adaptive parameters in the improved particle swarm optimization algorithm is improved, two disturbance methods of immune disturbance and chaos disturbance is introduced, and the annealing strategy and dynamic clustering strategy at the same time is applied. Experiments show that the algorithm has solved the above problem.

Key words: semi-supervised; *K*-means clustering; information entropy; perturbation; annealing strategy

随着应用环境复杂度的升高,传统 K 均值聚类 算法在处理含标称混合数据时[1],存在聚类中心过

收稿日期: 2020-02-16
基金项目: 河北省重点研发计划项目(20313701D); 河北省重点研发计划项目(19210404D); 国家自然科学基金项目(U1536112); 国家社会科学基金重点项目(17AJL014)
作者简介: 孙 艺(1979—), 男, 高级工程师, E-mail: sunyisse@bupt.edu.cn.

度敏感[2]、空聚类频繁出现[3],不能考虑类标号的作用[4-5]等问题,极大地限制了该算法处理数值和标称混合数据的能力. 对此,提出了一种改进的聚类方法:在传统粒子群算法基础上,改进了粒子群算法的自适应参数,引入了免疫扰动和混沌扰动2种扰动方式,应用了退火策略和动态聚类分析策略,减少了参数依赖,加快了粒子间的信息交换,进一步提高了全局搜索能力,保证了粒子种群的多样性,较好地避免了陷入局部极值情况的发生. 将自适应数目的初始聚类中心编码为粒子,再结合改进后的粒子群算法进行寻优,在聚类中引入标签信息作为软性约束,通过构建新的适应度函数,达到半监督聚类的效果. 通过实验证明,该算法在聚类质量和效果上有了很大提升.

1 K 均值聚类算法

1.1 K 均值聚类算法

K 均值聚类算法虽然具有诸多优点,但是中心选择的随机性,容易陷入局部最优,这方面的缺点比较突出,针对这些问题,较为成熟的方法是对模糊 K 均值聚类算法和 K 调和均值聚类算法等进行修改. 模糊 K 均值聚类算法的改进思路为在 K 均值算法的聚类过程中,为避免同一个样本点同属于2个簇的情况发生,引入模糊概念,在相似度评价函数中加入模糊因子,明确样本点在不同聚类中的隶属度,以代替算法中的相似度作为划分的标准. K 调和均值聚类算法通过改进聚类中心的选择方式,计算同簇内各样本点之间的平均距离,并用距离的调和平均值代替距离平均值,明显改善了初始化聚类中心敏感的问题.

1.2 粒子群算法

在粒子群算法中,采用向量编码方式表示粒子的状态及属性,位置表示各个维度的特征,速度表示位置在每次迭代中的变化量,对位置和速度进行迭代,最终得到最优解.

$$x_{i_d}(t+1)=x_{i_d}(t)+v_{i_d}(t+1) \tag{1}$$

$$v_{i_d}(t+1)=\omega v_{i_d}(t)+c_1r_1(t)[I_d(t)-x_{i_d}(t)]+c_2r_2(t)[G_d(t)-x_{i_d}(t)] \tag{2}$$

其中:d 为第 d 维的粒子属性,t 为当前迭代次数,$v_{i_d}(t)$ 为第 d 维上的控制粒子 i 在第 t 次迭代时的速度,c_1 和 c_2 均为常数(非负),表示社会学习因子和个体学习因子,ω 为惯性权重,$r_1(t)$ 和 $r_2(t)$ 为(0,1)之间的随机数,$I_d(t)$ 为粒子 i 具备最佳适应度的位置,$G_d(t)$ 为种群中所有粒子具备的最佳适应度的位置.

2 改进的 K 均值聚类算法

传统 K 均值算法在一定程度上解决了聚类中心敏感的问题,但仍存在易出现空聚类、无法处理数值、标称混合数据等问题. 因此,在调和均值算法与样本点隶属度2个因素并存情况下,在更新聚类中心时,计算所处样本点对该聚类中心的隶属度,以解决传统算法对初始聚类中心敏感度过高的问题,达到对模糊和不精确事实进行处理的目的;同时考虑标签信息对聚类结果的影响,引入软性约束项,即当聚类不满足软性约束时,在聚类目标函数计算时增加一个罚项或是补充项. 所以,新构建的具备半监督模糊性质的 K 调和均值目标函数由两部分组成,即评价聚类质量项和违反软性约束带来的罚项. 罚项由两部分组成,即违反 must-linked 关系的罚和违反 cannot-linked 关系的罚. 以此衡量用户的背景信息对聚类结果的影响程度,同时选用软性约束概念构建罚项,可以综合考虑距离和类标签对聚类结果的影响,对聚类中的坏点剔除也有一定帮助[1,6].

样本点 i 对聚类中心 j 的隶属度和聚类中心 l 更新公式分别为

$$\mu_{i,j}=\frac{\left(\frac{1}{d_{ij}^2}\right)^{\frac{1}{\alpha+1}}}{\sum_{j=1}^{K}\left(\frac{1}{d_{ij}^2}\right)^{\frac{1}{\alpha+1}}} \tag{3}$$

$$C_l=\frac{\sum_{i=1}^{N}\frac{\mu_{i,l}^{\alpha}}{\left(\sum_{j=1}^{K}\frac{\mu_{i,l}^{\alpha}d_{i,l}^2}{\mu_{i,j}^{\alpha}d_{i,j}^2}\right)^2}X_i}{\sum_{i=1}^{N}\frac{\mu_{i,l}^{\alpha}}{\left(\sum_{j=1}^{K}\frac{\mu_{i,l}^{\alpha}d_{i,l}^2}{\mu_{i,j}^{\alpha}d_{i,j}^2}\right)^2}} \tag{4}$$

其中:α 为隶属度控制因子,$\alpha\geqslant 0$ 时,$\sum_{j=1}^{K}\mu_{i,j}=1$,$\alpha=0$ 时,不考虑隶属度影响;X_i 为样本点;d_{il} 为样本点 i 到聚类中心 l 的欧氏距离的倒数;K 为聚类数;N 为数据点数目.

适应度函数公式为

$$F=\sum_{i=1}^{N}\frac{K}{\sum_{l=1}^{K}d_{i,j}^2\mu_{i,j}^{\alpha}}+\frac{N_{\text{label}}}{\lambda}\left(\sum_{\text{AB-must-linked}}\sum_{a=1}^{K}\sum_{b=a+1}^{K}\mu_{a,A}\mu_{b,B}\parallel C_a-C_b\parallel+\right.$$

$$\sum_{\text{AB-cannot-linked}} \sum_{a=1}^{K} \mu_{a,A}\mu_{a,B} \frac{\sum_{j,p=1}^{K} \| C_j - C_q \|}{K^2} \Bigg) \quad (5)$$

其中：N_{label} 为带有标签的样本点数量；λ 为自设权重，用于控制监督信息对聚类结果的影响程度；a，b 表示针对聚簇中心的遍历，A，B 表示针对数据点的遍历.

综上，提出一种新的方法，在最大程度上考虑用户的背景信息和数据本身信息，平衡用户背景知识与自适应的程度，挑选出最适合的数据集. 为避免空聚类问题的发生，采用式(6)对聚类数进行自适应计算，得出最佳聚类数 K.

$$\text{Fit}(K) = \frac{F}{\sum_{j,p=1}^{K} \| C_j - C_p \|} \quad (6)$$

其中 Fit 为聚类数目评价函数.

3　改进的粒子群算法

为解决传统粒子群算法存在诸多问题，引入粒子动态聚类策略，在算法的每次迭代中，将粒子划分为不同的聚类，以提高粒子间的信息交互效率. 粒子每次更新位置后都会重新进行聚类，使粒子可以在不同聚类中迁移，粒子间信息交换效率和算法的全局搜索能力被大大提高，多峰函数寻优能力不足和收敛速度较慢的缺陷也得到了极大改善[7].

3.1　动态聚类

使用调和 K 均值聚类算法对动态聚类策略进行了改进，有

$$C_l = \frac{\sum_{i=1}^{n} \frac{1}{\left(\sum_{j=1}^{K'} \frac{d_{il}^2}{d_{ij}^2} \right)^2} x_i}{\sum_{i=1}^{n} \frac{1}{\left(\sum_{j=1}^{K'} \frac{d_{il}^2}{d_{ij}^2} \right)^2}} \quad (7)$$

其中：K'为阈值，为$\frac{\sqrt{2}}{2}n$ 的聚类数；n 为总粒子数，x_i 为粒子位置. 样本点划分时采取硬聚类方法划入最近的聚类中心.

3.2　自适应参数

引入信息熵以结合自适应参数，代替随机数 r_2，以便更好地利用种群内其他粒子的信息，解决对惯性权重、学习因子过于依赖的问题. 当种群整体平均适应度大于或等于某一粒子聚簇平均适应度时，代表全局搜索能力的该粒子的惯性权重的取值也会随之提升；反之则降低该粒子的惯性权重，以增加其局部搜索能力，这就避免了传统算法在权重方面线性递减的缺陷.

3.2.1　惯性权重改进

传统算法中，惯性权重 ω 表示影响程度，表示上一次粒子的更新速度与本次粒子更新速度之间的关系. 笔者结合动态聚类策略，以自适应的方法控制惯性权重的变化，根据粒子聚簇的适应度和迭代次数，对惯性权重进行自适应判断，使之可以根据寻优过程中不同时期的搜索范围及能力的需求，自动调节搜索能力. 惯性权重自适应公式为

$$\omega_i(t+1) = \begin{cases} \omega_i(t) + [\omega_{\max} - \omega_i(t)]\, e^{-\frac{\bar{S}}{\bar{C}_j}}, & \bar{C}_j \geqslant \bar{S} \\ \omega_i(t) - [\omega_i(t) - \omega_{\min}]\, e^{-\frac{\bar{S}}{\bar{C}_j}}, & \bar{C}_j < \bar{S} \end{cases} \quad (8)$$

其中：$\bar{C}_j$ 为聚类中心 j 的平均适应度，$\bar{S}$ 为种群整体的平均适应度.

3.2.2　学习因子改进

在粒子群算法中，c_1 和 c_2 代表个体学习因子和社会学习因子，决定着粒子本身的经验信息和其他粒子的经验信息对粒子下一步运动的影响程度[8]. 为降低依赖关系，笔者采用一种根据迭代次数对社会因子和个体因子异步变化自适应的方法：在迭代过程中，2 种因子的取值随迭代次数的增加而线性减小，当固定的社会学习因子初始值为 2，个体学习因子初始值取 2.5，因子迭代终止值 0.5 时，算法效果最好. 如式(9)和式(10)所示.

$$c_1 = c_{1,\text{int}} + \frac{c_{1,\text{fin}} - c_{1,\text{int}}}{t_{\max}} t \quad (9)$$

$$c_2 = c_{2,\text{int}} + \frac{c_{2,\text{fin}} - c_{2,\text{int}}}{t_{\max}} t \quad (10)$$

其中：$t_{\max}$为预设的最大粒子群算法迭代次数，$c_{1,\text{int}}$和 $c_{1\text{fin}}$为预设的个体学习因子初始值和迭代终止值，$c_{2,\text{int}}$和 $c_{2,\text{fin}}$为预设的社会学习因子初始值和迭代终止值.

3.2.3　熵权引入

传统粒子群算法中，因为社会学习项选取的系数取值与随机数相结合，无法充分利用粒子信息，所以可引入熵权的概念，取代传统算法中粒子速度与社会学习因子相乘的随机数 r_2(见式(2))，充分利

用其他粒子的信息提高算法的搜索精度,提高收敛速度;经过 t 次迭代后,得到最优解 $n \times m$ 维矩阵 $\boldsymbol{D}$,m 为单个粒子维度数量,随后对最优解矩阵进行归一化处理,并计算各个维度的信息熵,为种群迭代 t 次后关于第 j 维粒子位置变化的信息熵

$$E_j = -\frac{1}{\ln n}\sum_{i=1}^{n} p_{ij}\ln p_{ij} \tag{11}$$

其中:p_{ij} 为概率,可由归一化矩阵计算得出,$j \in (1,2,\cdots,m)$,根据信息熵计算熵权信息为

$$\omega_j = \frac{E_j}{\sum_{i=1}^{n} E_i} \tag{12}$$

其中:ω_j 为所有粒子截止到第 t 次迭代时,个体最优解对第 j 维的熵权信息,用以替换传统粒子群中的社会学习项中的随机数.

3.3 扰动项改进

传统粒子群算法中,粒子种群最终会呈现出一种趋势,以个体极值和全局极值区间中的某位置为中心聚集,这就削弱了种群多样性和种群跳出局部最优的可能性. 为此,提出一种结合邻域扰动、免疫扰动和混沌扰动优点的方法,弥补上述不足.

3.3.1 邻域扰动改进

在传统算法中,因为过程不够优化,忽略了最好解的可能性;或因为各种因子选取不准确,降低了算法收敛速度等. 为此,结合邻域寻优的策略,对社会学习项的更新方式进行了改进,使得各粒子间可以进行直接的单向信息交流,利用各粒子的记忆性,记录较优解,沿梯度方向根据粒子速度决定搜索路径,保证了收敛速度,使搜索更新过程跟随当前最优解变化,有效地克服了搜索结果无明显改进的缺点.

改进后的速度更新公式为

$$v_{i_d}(t+1) = \omega(t)v_{i_d}(t) + c_1 r_1[I_d(t) - x_{i_d}(t)] + \frac{c_2 r_2}{2}\{[G_d(t) - x_{i_d}(t)] + [L_d - x_{i_d}(t)]\} \tag{13}$$

其中:L_d 为邻域最优解的第 d 维分量,$x_{i_d}(t)$ 为在第 t 次迭代时粒子 i 的 d 维分量取值.

3.3.2 免疫扰动改进

为解决粒子种群趋同性的问题,采用了免疫疫苗提取和免疫选择策略相结合的方法[9],选取当前平均适应度值最高的聚类,根据该聚类的聚类中心点和最大半径确定疫苗,解决以往算法中忽略粒子群体多样性造成的问题. 疫苗向量的第 d 维分量为

$$\boldsymbol{h}_d = \max\{|x_{1d} - e_d|, |x_{2d} - e_d| \cdots |x_{nd} - e_d|\} \tag{14}$$

其中:e 为适应度最优聚类的粒子平均位置. 通过代表粒子适应度和多样性的评价函数控制对粒子的疫苗接种操作. 评价函数为

$$P(x_i) = \alpha\frac{\sum_{j=1}^{n}|f(x_i) - f(x_j)|}{\sum_{i=1}^{n}\sum_{j=1}^{n}|f(x_i) - f(x_j)|} + \beta\frac{f(x_i)}{\sum_{i=1}^{n} f(x_i)} \tag{15}$$

α 和 β 为(0,1)的自设系数,函数前一部分为评价粒子浓度,后一部分为评价粒子的适应度. 由此可见,浓度较低的粒子和适应度较低的粒子更容易被选择接种疫苗. 因此,低浓度、高适应度的粒子被选择的概率较低;高浓度、低适应度的粒子被选择的概率也较低;而适应度较高,浓度也较高,即大量聚集于局部极值位置的粒子被选择的概率较大.

3.3.3 混沌扰动改进

通常免疫疫苗扰动只能保证被扰动粒子向当前全局最优位置移动的趋势,扰动的随机性和遍历性不强. 因此,引入自适应混沌扰动项,利用混沌理论的遍历性和随机性,优化扰动项. 混沌扰动项公式为

$$\boldsymbol{Q}_c(t) = \mu\boldsymbol{Q}_c(t-1)[1 - \boldsymbol{Q}_c(t-1)] \tag{16}$$

理论已经证明[10]:当参数 $\mu \in [0,4]$ 时,Logistic 混沌映射是[0,1]区间上的不可逆映射. 因此,结合免疫扰动和混沌扰动,笔者提出了粒子扰动处理的方法:

$$x_i'(t) = x_i(t) + \gamma_1 N(0,1)\boldsymbol{Q}_c(t) + \gamma_2 \boldsymbol{h}(t) \tag{17}$$

其中 γ_1 和 γ_2 为(0,1)之间的自设系数.

3.4 引入退火策略

引入模拟退火机制判断粒子早熟并及时使其跳出局部极值点,使用当前迭代时最佳粒子和最坏粒子的适应度差值调控温度,通过该差值的降低方式结合概率突跳特性,寻找目标函数的全局最优解,该方法不但延续了退火算法的性能,而且最大限度地保证了粒子群算法的收敛性. 粒子前后位置的适应度变化量为

$$\Delta = \frac{f_i(t+1) - f_i(t)}{f_{\text{worst}} - f_{\text{best}}} \tag{18}$$

接收概率为

$$P = \begin{cases} 1, & \Delta \leqslant 0 \\ e^{-\Delta}, & \Delta > 0 \end{cases} \tag{19}$$

其中:$f_i(t+1)$ 为粒子位置更新后的适应度;$f_i(t)$ 为粒子位置更新前的适应度;$f_{\text{worst}} - f_{\text{best}}$ 为当前迭代时

最好粒子和最坏粒子适应度的差值；Δ 为粒子前后位置的适应度变化量；$P > \text{random}[0,1]$ 时，粒子接受更新的位置，$P < \text{random}[0,1]$ 时拒绝更新位置；粒子变化时接受好解，但也可以以一定的概率接受坏解，避免陷入局部最优.

4　算法实现

算法分为半监督聚类部分和粒子群算法两部分，具体实现如下.

1）半监督聚类算法

① 根据需要聚类的数据点集合，随机选取 K 个元素作为数据中心.

② 根据式(4)更新聚类中心.

③ 根据式(6)评价当前的聚类数 K 是否恰当.

④ 重复① ~ ③步，每次迭代时聚类数 K 加 1，聚类数 K 的取值范围为 $2 \sim \frac{N}{\sqrt{2}}$，$N$ 为样本点数目.

2）粒子群算法

① 根据半监督聚类算法得到的聚类数 K，将粒子编码为 $K \times M$ 维向量，M 为数据点维数. 初始化粒子数目记为 n，$n \in [50,100]$. 根据适应度计算公式(13)计算个体最优和全局最优解.

② 根据式(7)对粒子进行动态聚类.

③ 根据式(12)和式(1)计算粒子新的位置.

④ 根据式(19)判断粒子位置是否更新.

⑤ 根据式(15)选择粒子并对其进行免疫扰动和混沌扰动处理.

⑥ 根据式(5)计算粒子适应度，与上次迭代的个体最优解和全局最优解进行对比，判断是否更新个体最优解和全局最优解.

⑦重复② ~ ⑥步，如果当前迭代次数达到预设值，退出算法.

5　仿真验证

将改进后的算法与原有经典算法对比发现，算法效果有较大提高，在标签数据占比为 50% 的数据集上，改进后的算法与原算法的比较结果如表 1 所示.

通过表 1 可以看出，改进后的算法在标签数据聚类的准确性和敏感性方面具有良好的表现，而且在聚类效果的各项评价指标要求方面都优于传统的 K 均值算法，改进后的粒子群算法性能比现有算法也有较大的提升.

表 1　聚类评价

性能评价指标	改进后算法	K-means
准确度	0.666 7	0.576 7
敏感性	0.837 5	0.937 5
特异性	0.604 5	0.445 5
精度	0.435 1	0.380 7
召回率	0.837 5	0.937 5
F-measure	0.572 6	0.541 5
gmean	0.711 6	0.646 2
Jaccard 系数	0.401 2	0.371 3

图 1 所示为 4 种粒子群算法的比较情况. 使用 Rosenbrock 函数为测试函数，图中 4 条曲线为目标函数值随算法迭代次数的变化曲线. 从图 1 可见，改进后的算法在收敛速度和降低对局部最优敏感度方面都要优于同种类型的算法，而且在精度方面的效果明显. 表 2 所示为当 Rosenbrock 函数迭代 100 次后，4 种算法在精度方面的对比情况，改进后的算法在精度方面明显提高了 2 ~ 4 个数量级.

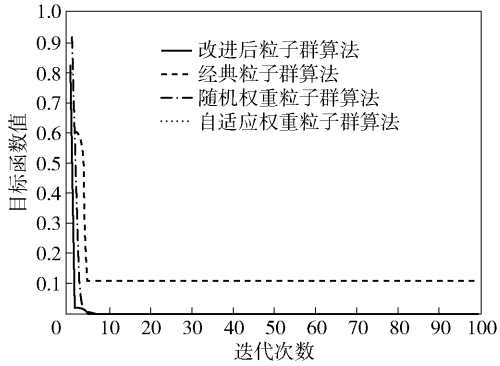

图 1　4 种粒子群算法的对比

表 2　不同粒子群算法的精度

算法	测试函数极小值数量级
改进后粒子群算法	$1\times10^{-24} \sim 1\times10^{-26}$
随机权重粒子群算法	$1\times10^{-19} \sim 1\times10^{-22}$
自适应权重粒子群算法	$1\times10^{-11} \sim 1\times10^{-13}$
经典粒子群算法	$1\times10^{-1} \sim 1\times10^{-2}$

当惯性权重自适应上限变为 1.2 和 0.9 时，改进后的算法与随机权重粒子群算法的性能比较如图 2 和图 3 所示.

可以看出，改进后的算法在不同参数情况下，其收敛速度和精度都提升了，算法在局部最优敏感度和参数依赖等方面都有较大的改善.

改进后半监督聚类算法的时间复杂度为

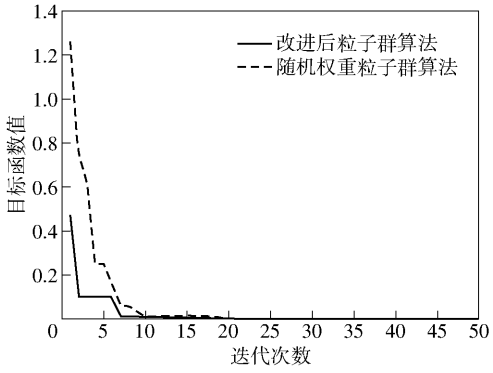

图2 改进后的算法与随机权重粒子群算法的比较

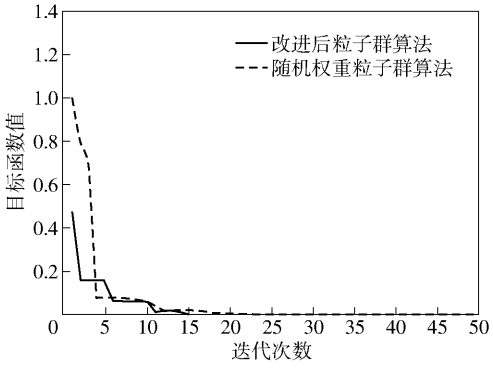

图3 改进后算法与随机权重粒子群算法的比较

$O(gkts)$,与经典 K 均值的时间复杂度 $O(gkt)$ 相比,不但考虑了诸多因素,而且复杂度并没有较大的变化,可见,算法整体效果良好. 其中 g 为数据点数量,k 为聚类中心数量,s 为监督信息种类数.

6 结束语

基于实现半监督 K 均值聚类目标,提出了一种改进的粒子群算法. 首先引入软性约束概念构造目标函数; 其次对粒子群算法进行改进和寻优,并对相关算法进行了验证. 通过实验数据分析,改进后的算法对标签数据处理的准确性和精度有了很大提升,在参数依赖性和局部最优敏感性方面有较大改进,对解决半监督聚类和寻优问题提供了一个良好的途径.

参考文献:

[1] Han Jiawei, Kamber Micheline. Data mining: concepts and techniques[J]. Data Mining Concepts Models Methods & Algorithms Second Edition, 2006, 5(4): 1-18.

[2] 刘佳鸣, 况立群, 尹洪红, 等. 灰狼优化的 K 均值聚类算法[J]. 中国科技论文, 2019, 14(7): 778-782, 807.
Liu Jiaming, Kuang Liqun, Yin Honghong, et al. K-means clustering algorithm based on grey wolf optimization [J]. China Sciencepaper, 2019, 14(7): 778-782, 807.

[3] 马俊宏, 武丽芬. 一种改进的加速 K 均值聚类算法[J]. 太赫兹科学与电子信息学报, 2019, 17(5): 885-891, 897.
Ma Junhong, Wu Lifen. An improved accelerated K-means clustering algorithm[J]. Journal of Terahertz Science and Electronic Information Technology, 2019, 17(5): 885-891, 897.

[4] Fu Mandi, Jian Yiheng, Yu Xiao, et al. Cross-company defect prediction via semi-supervised clustering-based data filtering and MSTrA-based transfer learning[J]. Soft Computing: A Fusion of Foundations, Methodologies and Applications, 2018, 22(10): 3461-3472.

[5] Yu Zhiwen, Luo Peinan, Liu Jiming, et al. Semi-supervised ensemble clustering based on selected constraint projection[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(12): 2394-2407.

[6] Yu Zhiwen, Chen Hongsheng, You Jane, et al. Double selection based semi-supervised clustering ensemble for tumor clustering from gene expression profiles[J]. Computational Biology & Bioinformatics IEEE/ACM Transactions on, 2014, 11(4): 727-740.

[7] 李文锋, 梁晓磊, 张煜, 等. 具有异构分簇的粒子群优化算法研究[J]. 电子学报, 2012, 40(11): 2194-2199.
Li Wenfeng, Liang Xiaolei, Zhang Yu, et al. Research on PSO with clusters and heterogeneity[J]. Acta Electronica Sinica, 2012, 40(11): 2194-2199.

[8] Doctor S, Venayagamoorthy G K, Gudise V G. Optimal PSO for collective robotic search applications[C]// Congress on Evolutionary Computation. Portland: IEEE, 2004.

[9] Sierra M R, Coello Coello C A. Improving PSO-based multi-objective optimization using crowding, mutation and E-Dominance[J]. Lecture Notes in Computer Science, 2005, 3410: 505-519.

[10] Liu Bo, Wang Ling, Jin Yihui, et al. Improved particle swarm optimization combined with chaos[J]. Chaos, Solitons and Fractals, 2005, 25(5): 1261-1271.