

改进的FCM算法及其应用^①

张忠厚, 赵 龙

(辽宁工程技术大学 理学院, 阜新 123000)

摘 要: 模糊C均值聚类算法是一种普遍应用的经典聚类算法, 在数据的分析方面有良好的表现, 但是算法的缺陷严重的限制了算法的应用和发展. 制药过程是一个十分复杂的综合系统, 被控参数情况也十分复杂, 因为有关联性和并且存在过失误差, 针对这些问题把模糊C均值算法应用到动态递归模糊神经网络预测控制当中. 利用改进的PSO算法对模糊C均值算法进行优化, 对数据的聚类辨识从而同步实现系统控制和异常预警, 保证系统稳定性.

关键词: 聚类; 粒子群优化算法; 模糊C均值; 隶属度

Improved FCM Algorithm and Its Application

ZHANG Zhong-Hou, ZHAO Long

(College of Science, Liaoning Technical University, Fuxin 123000, China)

Abstract: Fuzzy C-Means clustering algorithm is a classical clustering algorithm, which is widely used in data analysis and performs well. However, serious defects in the algorithm limit the application of the algorithm and development. Pharmaceutical process is a very complex integrated system, whose parameters are very complicated accused of the relevance and existence of fault and error. For these issues, the Fuzzy C-Means algorithm is applied to dynamic recurrent fuzzy neural network predictive control. Improve Fuzzy C-Means algorithm with IPSO algorithm to do the clustering of data and failure pattern recognition of system to achieve the early warning and ensure system stability.

Key words: clustering; particle swarm algorithm; fuzzy C-Means; membership

聚类方法有很多, 聚类分析指将物理或抽象对象的集合分组成为由类似的对象组成的多个类的分析过程. 它是一种重要的人类行为. 聚类分析的目标就是在相似的基础上收集数据来分类. 聚类源于很多领域, 包括数学, 计算机科学, 统计学, 生物学和经济学. 在不同的应用领域, 很多聚类技术都得到了发展, 这些技术方法被用作描述数据, 衡量不同数据源间的相似性, 以及把数据源分类到不同的簇中. 聚类分析是数据挖掘中的一个很活跃的研究领域, 并提出了许多聚类算法. 传统的聚类算法可以被分为五类: 划分方法、层次方法、基于密度方法、基于网格方法和基于模型方法. 其中FCM(Fuzzy C-Means)方法是应用最为广泛的一种聚类算法^[1].

模糊C均值聚类算法作为一个聚类的经典算法一直应用十分广泛, 虽然它存在的问题制约了它的发展, 为了保留和提高它的聚类能力, 近年来对其的研究和改进层出不穷, 它主要存在的问题主要在于两点: 一是算法的性能依赖于初始聚类中心的选取, 同时聚类的效果受初始值的影响较大; 二是FCM算法在迭代寻找最优解的过程中使用的是梯度下降的方法, 导致不可避免的会陷入局部最优值. 近年来对这两个问题而提出的改进方法有很多, 比如结合遗传算法的模糊聚类分析, 但针对需要更加高效的性能的系统来讲, 粒子群优化算法概念比较简单、实现容易, 搜索速度非常快. 所以粒子群优化算法应用的更加广泛. 在1996年, Krishnapuram 和 Keller 提出了PCM 算法(Possibi-

^① 收稿时间:2011-11-10;收到修改稿时间:2012-04-02

listic C-Means), 在原有的基础之上, 这种算法不仅包含了噪声数据的聚类而且使其对聚类结果影响较小, 但同时初始聚类中心敏感和易陷入局部收敛的问题并没有改进. 同 FCM 算法相比, 遗传算法、蚁群算法、粒子群算法对于初始值要求不高, 但由于他们自身的缺点, 很容易陷入局部最小值的问题. 但在这其中, 粒子群优化算法的高效性, 一直在很多领域有着较好的表现, 在预测控制当中, 时效性尤为重要, 数据分析的高效性直接影响着系统整个性能的高低, 所以在保持粒子群高效性的前提下解决它局部收敛的问题, 然后把它应用到模糊 C 均值的初始中心的确定上, 既保持了算法的高效性, 也优化了初始聚类中心, 提高了模糊 C 均值的性能^[2].

1 模糊C均值聚类

模糊 C 均值聚类方法的基本思想是: 将数据集 $X=\{x_1, x_2, \dots, x_n\} \in R^m$ 分为 C 类, X 中任意样本 x_k 对 i 类的隶属度为 μ_{ik} , 分类结果可以用一个模糊隶属度矩阵 $U=\{\mu_{ik}\} \in R^m$ 表示, 模糊 C 均值聚类是通过最小化关于隶属度矩阵 U 和聚类算法的迭代序列或其子序列必收敛到目标函数的一个极小值或者鞍点. 但这种算法本质上一种局部搜索算法, 容易陷入局部极小值. 如何为分类找到一个标准, 是需要解决的问题. 目前使用较多的是最小平方误差和^[3].

FCM 的目标函数值为:

$$J_m(U, E) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^m \|x_k - e_i\|^2 \quad (1)$$

m 是一个连续的数而且 $m > 1$. 聚类 i 表示为 e_i ($i=1, 2, \dots, c$). 样本 k 和聚类 i 间的隶属度表示为 μ_{ik} ($i=1, 2, \dots, c, k=1, 2, \dots, n$).

$$\mu_{ik} \in \{0, 1\}, \forall i, k; \sum_{i=1}^c \mu_{ik} = 1, \forall k \quad (2)$$

2 改进的模糊C均值聚类

2.1 PSO 算法

在标准 PSO 中, 粒子的当前位置与个体极值的位置并和全体极值的位置在不断的接近, 两个速度更新公式中的速度也在逐渐接近, 因此, 在更新粒子的速度在迭代后期主要是根据 $\omega \times v$ 变化的. 当惯性权重线性下降到小于 1 时, 粒子的飞行速度逐渐变小, 最后接近于零. 因为没有飞行速度, 粒子出现“惰性”, 随

着迭代的进行, 其他粒子将很快向惰性粒子周围聚集, 发生“收敛”的现象, 这使得算法出现过早收敛到局部最优^[4-6]. 由于它的高效和易于实现, 很多学者希望在保留它的优势的前提下改进它过早收敛于局部最优, 对于标准 PSO 算法的改进已经有很多方法, 融合其他算法会增加算法的复杂度, 是以速度为代价的, 而且不易于实现. 对于参数的改进也有很多, 有增加了一些参数来控制局部收敛, 但这无疑增加了算法的复杂性, 而对于原有的参数的特定设置又增加了人为因素, 这和增加算法和系统的自控能力不符, 所以对于 PSO 算法的改进依然是个研究热点. 本文采用的是结合 Suganthan 的带有邻域操作的 PSO 模型进一步改进更新策略, 避免粒子过早的局部收敛.

更新策略:

$$v(t+1) = \omega v(t) + c_1 r_1 (p_i - x(t)) + c_2 r_2 (p_g - x(t)) \quad (4)$$

$$x(t+1) = x(t) + v(t+1) \quad (5)$$

其中, ω 是惯性权重; p_i 代表个体极值; p_g 代表全局极值; c_1 和 c_2 分别是个体和全局加速系数; r_1 和 r_2 分别是 0 和 1 之间的随机数; 为了使 PSO 收敛到全局最优同时保证收敛速度, 结合 Suganthan 的带有邻域操作的 PSO 模型定义粒子 s_i 的邻域极值为 l_i , 将该极值也作为粒子进化的一个信息来源^[7-8]. 在优化的初始阶段, 将邻域定义为每个粒子自身, 随着迭代次数的增加, 将邻域范围逐步扩展到包含所有粒子这样就避免了上述情况出现而导致早熟收敛, 新的更新策略调整为

$$v_i = \omega v_i + c_1 r_1 (p_i - s_i) + c_2 r_2 (g - s_i) + c_3 r_3 (l_i - s_i) \quad (6)$$

算法具体实现步骤:

步骤 1: 初始化粒子群和模糊聚类的参数, 包括群体规模, 粒子的初始位置, 粒子的初始速度, 样本数, 聚类中心和模糊指数等; 设置 ω , c_1 和 c_2 .

步骤 2: 根据初始位置值计算出相应的聚类中心, 然后计算目标函数值和相应的聚类中心, 存储每个粒子的最佳位置和最佳适应值.

步骤 3: 根据公式更新每个粒子的速度和位置.

步骤 4: 用约束的策略, 使粒子的值满足模糊聚类的制约. 然后根据粒子的值计算出相应的聚类中心, 根据粒子和聚类中心, 计算出每个粒子的适应值, 比较适应值.

步骤 5: 比较每个粒子当前和以前的最佳适应值, 如果当前更优, 则设置当前最有位置为最优. 比较当

前最优值和全局最优值,更新全局最优值.

步骤 6: 搜索算法, 确定是否满足结束条件, 如果不满足, 则返回步骤 3, 满足则停止迭代, 输出最优位置.

3 数据辨识

改进后的 FCM 算法主要是应用于过程控制中的数据辨识, 在前期的数据处理过程中, 对数据进行筛选和分类, 有利于在过程控制时的及时提供有效信息, 大大的提高了系统的性能. 在这里主要介绍的是只要系统中的应用. 结合 DRFNN 和 PSO 算法形成制药系统中的预测控制的部分. DRFNN 同时具有递归神经网络和模糊逻辑特点, 它通过与外部的延时反馈可以较好的反映出系统的动态映射关系, 因此比前馈神经网络更具有优势, 而模糊逻辑具有定性知识表达的能力, 所以 DRFNN 在处理参数漂移、强干扰、非线性、不确定性等问题上表现出了较好的性能. DRFNN 采用五层结构: 除输入层、输出层外, 第二层为规则的前件层, 这一层的每个节点代表输入变量的一个值, 采用高斯函数进行模糊化处理. 第三层为规则层, 进行模糊推理, 第四层为状态层, 其节点的输出为从规则层获得的推理结果的线性组合, 还包括一个内部的反馈环, 利用它捕获系统的动态响应而不必在外部添加延时反馈环节^[9].

动态递归模糊神经网络结构复杂, 而粒子群算法可以优化全连接网络结构下的各层之间的连接权值和优化网络的拓扑结构. 改进的粒子群优化算法也很大程度上解决了早熟收敛的问题, 有很好的泛化能力, 在实际应用中改进的粒子群优化算法原理更简单, 参数更少, 实现更容易^[10]. 应用了改进的 PSO 算法的 DRFNN 有更好的预测性, 使控制有更好的动态性、鲁棒性和稳定性. 这种控制系统不但响应的速度快, 超调量小, 并且具有波动幅度小、运行平稳的特点, 满足了制药过程中对跟随给定序列和限制系统超调的要求, 体现了良好的功能特性. 把改进后的模糊 C 均值聚类算法和神经网络结合, 对系统的实时数据进行分析聚类识别, 在预测控制的同时提取异常数据进行异常识别.

4 仿真

本文分别使用 FCM PSO-FCM 和 IPSO-FCM 对 UCI 数据库中的 Iris 和 Wine 数据集来验证改进的聚类

算法的聚类效果并把算法应用到系统中对青霉素的菌体浓度做了预测. 仿真实验使用 Matlab2008a 进行仿真. 而针对预测控制的实验主要集中在预测值和实际值的偏差来体现模糊神经网络预测控制的性能和改进后的模糊 C 均值在系统中的应用效果. 制药系统有一定的相似性, 属于过程控制的一部分, 一般都包括温度, 压力, 密度等, 各个参数都有着关联性.

表 1 可以看出, 基于改进粒子群算法的模糊 C 均值算法在所有指标上有优于其它算法. 其中, 传统算法由于是一种梯度下降算法, 函数值下降的非常迅速, 非常容易陷入局部最优, 方差较大说明最优解的不稳定性, 对初值敏感, 基于 PSO 算法的 FCM 算法在初值敏感的问题上有改进, 但在局部收敛的问题上, 并没有取得大的进步, 改进后的 PSO 算法较 PSO 算法有更好的全局搜索能力, 所以 IPSO-FCM 在各项性能都优于其它算法^[11,12].

表 1 基于数据集 iris 和 wine 的聚类效果

数据集	算法	J _m (U, W)		正确率 %	迭代数
		均值	方差		
Iris	FCM	80	8.46	88.7	21
	PSO-FCM	60	0.46	92.0	100
	IPSO-FCM	44	0.43	93.4	100
Wine	FCM	618	70.1	69.1	46
	PSO-FCM	584	17.3	76.4	100
	IPSO-FCM	563	12.2	79.8	100

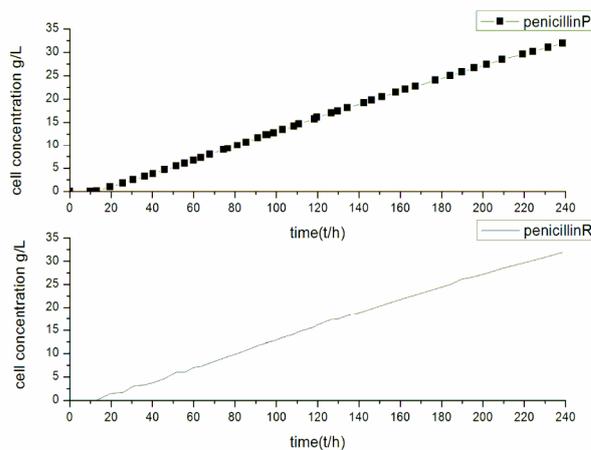


图 1 青霉素菌体浓度预测结果比对图

图 1 可以看出改进的模糊 C 均值算法在系统中的

应用效果良好,可以应用到实际的生产当中.结合该方法的系统具有一定的高效性和稳定性,有较高的预测性和实用性.

5 结语

本文提出了一种基于改进的 PSO 算法的模糊 C 均值聚类算法,把这种算法应用到了制药系统的控制当中,在系统已经提高了鲁棒性和高效性的基础之上,加入来故障识别,从而保证了系统的安全性.改进的 PSO 算法不仅克服了局部收敛的问题还很大程度上保持了算法的高效性.实验结果表明这种改进机制是可行的,把改进后的 FCM 算法应用到制药系统当中也是可行的并且发挥了很好的作用.

参考文献

- 1 薛尧予,王建林,于涛,赵利强.基于改进 PSO 算法的过程模型参数估计.仪器仪表学报,2010,31(1):178-182.
- 2 Lakshmipathi SR. Implementation of an Innovative Bio Inspired GA and PSO Algorithm for Controller design considering Steam GT Dynamics. Journal of Computer Science, January 2010,24(2):132-140.
- 3 武俊峰,艾岭.一种基于改进算法的模糊模型辨识.哈尔滨理工大学学报,2010,20(5)165-170.
- 4 穆朝絮,张瑞民,孙长银.基于粒子群优化的非线性系统最小二乘支持向量机预测控制方法.控制理论与应用,2010, 27(2):164-168.
- 5 王博,孙玉坤.基于数据场聚类的模糊神经网络在发酵过程中的应用.仪器仪表学报,2009,30(5):944-948.
- 6 Xue YY, Wang JL, Yu T, Zhao LQ. Parameter estimation of fermentation process model based on an improved PSO algorithm. Chinese Journal of Scientific Instrument, 2010, 31(1):178-182.
- 7 田雨.混合神经网络技术.第 2 版.北京:科学出版社, 2009. 148-160.
- 8 吕奕清,林锦贤.基于 MPI 的并行 PSO 混合 K 均值聚类算法.计算机应用,2011,32(1)112-118.
- 9 李国勇.智能控制及其 MTLAB 的实现.第 3 版.北京:电子工业出版社,2007.156-188.

(上接第 94 页)

行为分析.所以,接下来的工作是寻找更佳的方法验证稀疏对象的视频快速浏览的有效性.

参考文献

- 1 Chalidabhongse TH, Kim K, Harwood D, et al. A perturbation method for evaluating background subtraction algorithms. Proc. of the Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. Nice, France: 2003. 11-12.
- 2 Hall D, Nascimento J, Ribeiro P, et al. Comparison of target detection algorithms using adaptive background models. 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. Beijing, China, 2005. 113-120.
- 3 王春涛.基于背景差分法和光流法的视频动态目标检测与跟踪.软件导刊,2011,10(6):145-147.
- 4 Stauffer C, Grimson WEL. Learning patterns of activity using real-time tracking. Proc IEEE Trans on PAMI. Washington, 2000. 747-757.
- 5 Poppe C, Martens G, Lambert P, et al. Mixture models based background subtraction for video surveillance applications. Proc. of the 12th International Conference on Computer Analysis of Images and Patterns. Vienna, Austria, 2007: 28-35.
- 6 Majdi-Nasab N, Analoui M. Decomposing parameters of mixture Gaussian model using genetic and maximum likelihood algorithms on dental images. Pattern Recognition Letters, 2006,27(13):1522-1536.
- 7 Perez P, Gangnet M, Blake A. Poisson image editing. Proc. of ACM SIGGRAPH. 2003. 313-318.
- 8 Jia JY, Sun J, Tang CK. Drag-and-drop pasting. Proc. of ACM SIGGRAPH. 2006. 631-637.
- 9 宋雪华,陈瑜,耿剑锋,陈景柱.基于改进的混合高斯模型的运动目标检测.计算机工程与设计,2010,31(21):4646-4649.