

基于粒子群聚类算法的大坝安全监控模型

王 伟¹, 沈振中¹, 王连庆²

(1. 河海大学水利水电工程学院, 江苏 南京 210098; 2. 江苏弘盛建设工程集团有限公司, 江苏 高邮 225600)

摘要 将粒子群算法与模糊聚类算法相结合, 建立了基于粒子群聚类算法的大坝安全监控模型. 该算法将分类矩阵作为粒子的编码形式, 依据粒子的个体极值和全局极值, 充分利用正反馈计算信息, 自适应性地确定模糊分类矩阵和聚类中心. 工程算例表明: 粒子群聚类算法进一步提高了聚类算法的区间预报能力. 对于高维优化问题, 粒子的搜索过程比较复杂, 该算法的收敛速度较慢.

关键词 粒子群算法; 模糊聚类; 大坝; 安全监控模型

中图分类号: TV698 文献标识码: A 文章编号: 1000-198X(2008)04-0501-04

大坝安全是关系到国计民生的大事, 如何有效挖掘监测数据的内在信息, 一直是国内外学者研究的热点. 由于大坝安全监测的多样性和复杂性, 效应量(变形等)与影响量(水位和温度等)存在大量的不确定性. 这种不确定性具有随机性和模糊性, 应用模糊数学能够较好地反映大坝变形的模糊性^[1]. 在大坝安全监控方面, 主要应用模糊数学中的聚类分析法进行研究. 如: 王绍泉^[2]提出了大坝安全分析的多层次阈值模糊综合评判模型, 徐洪钟等^[3]提出了一种应用于大坝安全监控的自适应模糊神经网络, 蔡新等^[4]将模糊理论引入土石坝优化设计, 冯福恒等^[5]依据模糊控制理论, 应用模糊聚类分析法建立了复杂结构混凝土坝变形的预测模型, 王铁生等^[6]将模糊聚类算法与神经网络相结合, 提出了基于模糊神经网络的大坝监控模型.

粒子群优化算法(简称 PSO)由 Kennedy 等^[7]提出. 该算法的基本思想是: 通过对局部极值、全局极值的“记忆”来搜寻最优解. 相比其他仿生算法, 该算法具有调整参数少、收敛快、简单易实现等优点. 本文利用粒子群算法的全局搜索能力确定模糊聚类算法中的分类矩阵, 并建立相应的大坝安全监控模型, 通过工程算例验证该模型的应用可行性.

1 模糊聚类算法

设待分类的样本个数为 n , 每个样本 x 均有 m 个特征指标. 如果把样本集分成 c 类, 可用一个 c 行 n 列、元素 a_{ik} 在 $[0, 1]$ 之间的矩阵 A 来表示每个样本以某一从属程度从属于某一类, 而又以另一从属度从属于另一类. 因此, 矩阵 A 具有如下 3 个特性:

$$0 \leq a_{ik} \leq 1 \quad \sum_{k=1}^n a_{ik} > 0 \quad \sum_{i=1}^c a_{ik} = 1 \quad (i = 1, 2, \dots, c; k = 1, 2, \dots, n) \quad (1)$$

模糊聚类算法的基本步骤为: 选取初始模糊分类矩阵 $A^{(0)}$, 对于当前的分类矩阵, 以式(2)计算各聚类中心矩阵 $V^{(l)}$. 通过式(3)依据新的聚类中心修正模糊分类矩阵 $A^{(l+1)}$. 给定差值精度 ϵ , 比较迭代前后的分类矩阵的差值, 若满足式(4), 则停止迭代, $A^{(l+1)}$, $V^{(l)}$ 即为所求的分类矩阵和聚类中心, 否则继续迭代直至满足差值精度要求.

$$v_i^{(l)} = \frac{\sum_{k=1}^n (a_{ik}^{(l)}) x_k}{\sum_{k=1}^n (a_{ik}^{(l)})} \quad (i = 1, 2, \dots, c) \quad (2)$$

$$a_{ik}^{(l+1)} = \left[\sum_{j=1}^c \left(\frac{\|x_k - v_i^{(l)}\|}{\|x_k - v_j^{(l)}\|} \right)^{2(r-1)} \right]^{-1} \quad (k = 1, 2, \dots, m) \quad (3)$$

$$\max\{|a_{ik}^{(l+1)} - a_{ik}^{(l)}|\} \leq \varepsilon \quad (i = 1, 2, \dots, c; k = 1, 2, \dots, m) \quad (4)$$

式中: v_i ——第 i 类的聚类中心; r ——加权指数, 又称为平滑因子; l ——迭代次数。

2 基于粒子群聚类算法的大坝安全监控模型

用聚类算法作为大坝安全监控预报模型就是将大坝安全监控值(效应量)按照规则划分为几个值区域, 先对部分监控数据进行聚类计算, 将计算所得到的聚类中心作为基础, 再对待测监控数据进行二次聚类分析, 从而确定待测数据属于哪个值区域, 以达到区间预报的目的。由于粒子群属于随机优化算法, 因此要将模糊聚类算法中的计算条件转换为优化问题, 以便于采用粒子群算法进行优化求解。可将条件(4)转换为求解最小值优化问题, 优化目标定义为样本到聚类中心的距离平方和为最小, 即

$$\min(f(k)) = \sum_{j=1}^n \sum_{i=1}^c (a_{ij}) \|x_j - v_i\|^2 \quad (k = 1, 2, \dots, N) \quad (5)$$

式中: $f(k)$ ——目标函数; N ——候选解的组数。

确定粒子的编码形式是应用粒子群算法的关键, 不同的优化问题有不同的粒子编码形式。已有文献^[8]常将聚类中心向量作为粒子, 但若将分类矩阵作为粒子编码形式, 通过迭代可同时求得最优的分类矩阵和聚类中心。粒子群中的粒子被看作为一个没有质量和体积的微粒, 由位置和速度 2 个向量决定其性质。若监控数据有 n 组, 分类类型有 c 种, 粒子种群数为 N , 则每个粒子的位置编码为 $c \times n$ 维矩阵数组, 相应的速度编码也为 $c \times n$ 维矩阵数组:

$$V_k = (V_{ij}) \quad x_k = (a_{ij}) \quad (i = 1, 2, \dots, c; j = 1, 2, \dots, m; k = 1, 2, \dots, N) \quad (6)$$

每个粒子在 n 维搜索空间飞行时, 通过式(7)更新自己的移动速度。

$$V^{(l+1)} = wV^{(l)} + c_1 r_a (p^{(l)} - x^{(l)}) + c_2 r_a (g^{(l)} - x^{(l)}) \quad (7)$$

式中: $V^{(l+1)}$ ——粒子当前的速度向量; $V^{(l)}$ ——粒子的历史速度向量; $p^{(l)}$ ——粒子个体极值所对应的位置向量; $x^{(l)}$ ——粒子的历史位置向量; $g^{(l)}$ ——粒子全局极值所对应的位置向量; w ——惯性权重; c_1, c_2 ——学习因子; r_a ——(0, 1)之间的随机数。

惯性权重和学习因子均为可以调整的常数。惯性权重是反映粒子历史速度对当前速度的影响程度的一个量, 也是影响粒子全局或局部搜索能力的一个量。计算时常通过线性递减变化的手段来选用惯性权重^[9-10]。因为在计算初期, 粒子的搜索具有随机性, 采用较大的惯性权重可提高粒子的全局搜索能力, 加大粒子的搜索范围, 随着迭代过程的进行, 历史速度对粒子的影响逐步削弱, 算法的收敛速度进一步加快。式(7)等号右边的第 2 项表示粒子在历次搜索过程中, 向自身的最优解方向飞行; 第 3 项表示粒子向粒子种群中最优解的方向飞行。由文献[9]可知, 惯性权重 w 一般在(0, 1)之间取值, 学习因子 c_1, c_2 常取相同的常数, 一般为 $c_1 = c_2 = 2$ 。粒子的更新位置由移动速度决定, 更新位置公式为

$$x^{(l+1)} = x^{(l)} + V^{(l+1)} \quad (8)$$

式中: $x^{(l+1)}$ ——粒子的当前位置向量。另外, 将每次迭代过程中粒子的速度限制在 $[V_{\min}, V_{\max}]$ 之间, 以保证粒子群优化算法的收敛性。

粒子适应度与算法中的目标函数相关。在该算法中, 目标函数定义为样本与聚类中心的距离平方和, 将式(5)作为优化目标, 相应的粒子适应度为

$$F(k) = \sum_{j=1}^n \sum_{i=1}^c a_{ij}^r \|x_j - v_i\|^2 \quad (i = 1, 2, \dots, c; j = 1, 2, \dots, m; k = 1, 2, \dots, N) \quad (9)$$

由式(9)可以看出, 样本与聚类中心的距离平方和越小, 粒子的适应度就越小, 即粒子的位置向量就越逼近最优解。式(7)~(9)是粒子群聚类算法的核心部分, 在历次迭代过程中, 各粒子依据其适应度, 通过对个体极值、全局极值的“记忆”来更新其位置和速度。

基于粒子群聚类算法的计算步骤如下:

a. 设置粒子群各参数和最大迭代数, 假定粒子种群数为 N , 每个粒子的位置向量代表一组大坝安全监控模型的初始分类矩阵的候选解, 且随机所生成的初始分类矩阵中元素应符合式(1)的 3 个特征。

b. 依据式 (9) 计算各粒子的适应度 $F(k)$, 并将 $F(k)$ 作为各粒子的初始个体极值, 比较各个体极值求出初始全局极值.

c. 依据式 (7) (8) 更新各粒子的速度和位置.

d. 重新计算各粒子的聚类中心向量和当前适应度 $F^{(l+1)}(k)$, 并与该位置下的历史个体极值 $p^{(l)}(k)$ 相比较, 若 $F^{(l+1)}(k) < p^{(l)}(k)$ 则将 $F^{(l+1)}(k)$ 作为粒子的个体极值. 再次比较各粒子的个体极值, 找出粒子群的全局极值.

e. 重复步骤 c, d, 直至达到最大迭代数. 输出粒子群最优全局极值情况下的聚类中心向量.

f. 计算新样本与聚类中心的距离, 以最短距离划分样本的归类, 输出新样本的预报区间.

根据以上计算步骤, 采用 VB 语言编译了基于粒子群聚类算法的大坝安全监控模型的计算程序, 用于确定样本分类矩阵, 并对坝体变形进行区间预报.

3 工程算例

新安江水电站的拦河坝为混凝土宽缝重力坝, 坝顶高程 115.0m, 最大坝高 105.0m, 坝顶全长 466.5m, 由 26 个坝段组成, 编号从 0~25. 大坝坝顶水平位移的观测采用视准线法, 从右至左 2~20 号、22 号和 23 号坝段共有 21 个水平位移测点, 分别由 3 条视准线进行观测.

以 10 号坝段的 1980~1990 年之间的水平位移作为计算数据, 共 140 组. 从成因分析, 混凝土坝的位移主要受水压分量、温度分量、时效分量的影响. 其中: 水压因子(上下游水位差)取 3 个, 即 H_1, H_2 和 H_3 ; 温度因子取 6 个, 即 T_1 (当天气温), T_5 (前 5d 平均气温), T_{10} (前 10d 平均气温), T_{30} (前 30d 平均气温), T_{60} (前 60d 平均气温)和 T_{90} (前 90d 平均气温); 时效因子取 2 个, θ 和 $\ln\theta$ (θ 为观测日至基准日的累计天数除以 100). 因为各因子具有不同的量纲, 需通过规格化处理来消除量纲的影响. 仿真试验比较表明, 选用极差规格化较为合理. 对 140 组监测数据的统计分析表明, 标准差 s 在 4.09mm 左右, 取置信度 95%, 则带宽 $\Delta = 1.96s$. 因此, 水平位移分区间隔取为 8mm. 考虑到水平位移 $L_{max} = 22.19$ mm, $L_{min} = 5.06$ mm, 将水平位移分为 [5, 13], [13, 21] 和 [21, 29] 三个区, 分别对应 I, II 和 III 3 个程度等级.

以前 100 组监测数据作为训练样本, 后 40 组监测数据作为待测样本, 计算参数设置为: 粒子群数为 50, 迭代 500 次, 惯性权重从 0.9 递减到 0.4, 学习因子 $c_1 = c_2 = 2$. 聚类效果的检验采用分类系数和平均模糊熵, 分类系数越接近 1, 最终分类的模糊性越小, 聚类效果就越好; 平均模糊熵越接近 0, 其聚类效果就越好. 模糊聚类算法和粒子群聚类算法的聚类效果对比如表 1 所示.

表 1 聚类效果的检验指标对比

Table 1 Comparison of clustering effects of the fuzzy clustering algorithm and the PSO-fuzzy clustering algorithm

计算方法	分类系数	平均模糊熵
模糊聚类算法	0.944	0.0938
粒子群聚类算法	0.979	0.0388

由表 1 可知, 粒子群聚类算法的分类系数值比模糊聚类算法更接近 1、平均模糊熵值更接近 0, 所以粒子群聚类算法的聚类效果与模糊聚类算法相比较优. 以计算所得的样本聚类中心为基础, 对后 40 组数据进行区间预报. 结果表明, 粒子群聚类算法的区间预报正确率为 75%, 模糊聚类算法的区间预报正确率为 57.5%. 因此, 采用基于粒子群聚类算法的大坝安全监控模型进行预报是可行的.

因为影响因子数目较多, 所以应用粒子群聚类算法求解聚类问题, 属于高维度优化问题. 在计算过程中, 其目标函数值随迭代次数变化的收敛过程如图 1 所示, 当计算迭代次数达到 300 后, 目标函数值变化才趋于稳定, 因此该算法的收敛速度较慢.

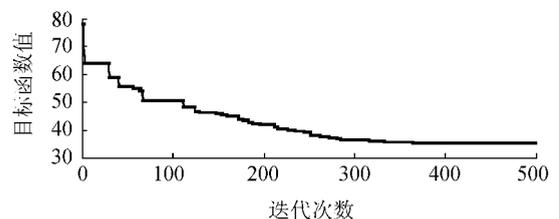


图 1 目标函数值的收敛过程

Fig.1 Objective function value versus iteration number for the PSO-fuzzy clustering algorithm

4 结 语

a. 粒子群算法是目前应用前景较为广阔的仿生智能群算法之一. 虽然模糊聚类算法与粒子群算法都属于目标函数最优化的无监督学习算法, 但后者依据局部极值及全局极值不断更新其速度和位置, 充分利用正反馈信息, 动态地逼近最优解.

b. 工程算例表明,利用粒子群算法确定模糊分类矩阵,使目标函数值为最小的计算过程更加优化.基于粒子群聚类算法的样本聚类效果与模糊聚类算法相比较优,区间预报精度有所提高.因此,将基于该算法的大坝安全监控模型用于预报是可行的.

c. 对于高维度空间优化问题,因为影响大坝变形量的因子数目较多,粒子在高维空间搜索最优解的过程比较复杂,所以粒子群聚类算法的收敛速度较慢.如何提高粒子群聚类算法的收敛速度和计算精度是需要进一步研究的问题.

参考文献:

- [1] 吴中如,沈长松,阮焕祥.水工建筑物安全监控理论及其应用[M].南京:河海大学出版社,1990.
- [2] 王绍泉.多层次阈值模糊综合评判在分析大坝安全中的应用[J].大坝观测与土工测试,1997,21(4):12-14.
- [3] 徐洪钟,胡群革,吴中如.自适应模糊神经网络在大坝安全监控中的应用[J].河海大学学报:自然科学版,2001,29(2):8-10.
- [4] 蔡新,杨建贵,王海祥.土石坝广义模糊优化设计[J].河海大学学报:自然科学版,2001,29(1):24-27.
- [5] 马福恒,王仁钟,吴忠如,等.模糊控制的预测模型及其应用[J].大坝观测与土工测试,2001,25(2):17-20.
- [6] 王铁生,华锡生.基于模糊聚类算法的大坝监控模型的研究[J].水利学报,2003(6):115-118.
- [7] KEMEDY J, EBERHART R C. Particle swarm optimization[C]//Proceedings of IEEE International Conference on Neural Networks. Piscataway, NJ:IEEE Service Center,1995:1942-1948.
- [8] 高尚,杨静宇.一种新的基于粒子群算法的聚类方法[J].南京航空航天大学学报,2006,38(增刊):62-65.
- [9] SHI Yu-hui, EBERHART R C. Parameter selection in particle swarm optimization[J]. Evolutionary Programming VII, Spring Berlin, 1998, 1447:591-600.
- [10] 周驰,高海兵,高亮,等.粒子群优化算法[J].计算机应用研究,2003(12):7-11.

Dam safety monitoring model based on PSO-fuzzy clustering algorithm

WANG Wei¹, SHEN Zhen-zhong¹, WANG Lian-qing²

(1. College of Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China;

2. Jiangsu Hongsheng Construction Engineering Group Co., LTD, Gaoyou 225600, China)

Abstract: A dam safety monitoring model was presented in this paper by combining a fuzzy clustering algorithm and a particle swarm optimization (PSO) algorithm. Based on the position vector of particles represented by a classification matrix, the individual extremum and global extremum of each particle and the positive feedback information in the PSO, the fuzzy classification matrix and clustering center were adaptively determined. The result of an engineering application shows that, compared with the traditional fuzzy clustering algorithm, the PSO-fuzzy clustering algorithm improves the clustering effect and interval forecasting ability. It is also concluded that for high-dimension optimization problems, the convergence speed of this algorithm is slow because of its complex search process.

Key words: particle swarm optimization; fuzzy clustering; dam; safety monitoring model