

# 电力系统粒子群优化模糊聚类算法及其应用

何晓峰, 王 钢, 李海锋

(华南理工大学电力学院, 广东 广州 510640)

**摘要:** 电力系统数据信息处理与应用领域广泛涉及数据挖掘与特征提取问题, 为了提高其中聚类算法的有效性, 提出了一种粒子群优化算法(PSO)与模糊C均值算法(FCM)有机结合的粒子群优化模糊聚类算法。该算法用 PSO 优化过程代替 FCM 中的基于梯度下降的迭代过程, 充分利用 PSO 具有全局寻优、快速收敛的特点, 使算法具有很强的全局搜索能力, 有效地避免了 FCM 易陷入局部极小的缺陷; 同时也降低了 FCM 对初始值的敏感度。还通过核方法, 将低维特征空间的样本通过核函数映射到高维特征空间, 增强了特征的优化, 使特征在高维空间更易聚类。电力系统负荷样本聚类的应用仿真研究结果表明: 与单纯 FCM 法相比, 该算法聚类更准确, 效果更佳。

**关键词:** 模糊聚类; 数据挖掘; 粒子群优化; 负荷样本

## Power system clustering algorithm combining particle swarm optimization with fuzzy C means and its application

HE Xiao-feng, WANG Gang, LI Hai-feng

(College of Electrical Engineering, South China University of Technology, Guangzhou 510640, China)

**Abstract:** Information process and its application of power system relate widely to the technology of data mind and character extraction. In order to increase the efficiency of clustering algorithm, a novel clustering algorithm of power system, which combines particle swarm optimization (PSO) with fuzzy C means (FCM) organically, is presented. The algorithm uses PSO to replace the iteration process of FCM based on the gradient descent. So it has strong global searching capacity, and avoids the problems of local optimization and initialization values. At the same time, the kernel method is involved to map the eigenspace of low dimensions to that of high dimensions by kernel function, so the optimization of eigenvalue is enhanced, and the eigenvalue is easier to be clustered in the eigenspace of high dimension. The simulation result of power system load clustering shows that the proposed algorithm is more accurate and efficient than FCM.

**Key words:** fuzzy clustering; data mining; particle swarm optimization; load sample

中图分类号: TM715 文献标识码: A 文章编号: 1003-4897(2007)22-0040-05

## 0 引言

数据挖掘的任务就是从数据集中发现模式, 而模式种类繁多, 在实际应用中往往根据模式的实际作用将数据挖掘分为以下几种: 分类、聚类、关联、序列等。聚类就是将数据对象分组成为多个类或簇, 在同一个类/簇中的对象之间具有较高的相似度, 而不同类/簇中的对象差别较大。聚类分析可以应用于电力系统的多个领域, 如: 数据预处理、数据分析、相似日选择、负荷特征提取、电网故障诊断、变压器故障诊断、发电机故障诊断、电能质量评估、负荷特性分类和电力系统稳定状态分类等。目前该领域的研究已有不少成果, 如: 文献[1]提出了基于模糊C均值法的负荷特性分类与综合; 文献[2]提出了基于蚁群优化算法的电力系统负荷序列聚类分析;

文献[3]提出了一种基于粒子群优化的并行神经网络的电力系统负荷聚类算法。

模糊C均值 FCM (Fuzzy C Means) 聚类算法是聚类算法应用最广泛的一种。由于 FCM 算法本质是用梯度下降的方法寻找最优解, 因此存在局部最优问题; 同时算法的收敛速度受初始值的影响也较大, 特别是在聚类数较大的情况下, 这一缺点更为突出。粒子群优化 PSO (Particle Swarm Optimization) 算法是一种随机优化技术, 在许多领域得到成功应用, 并且在多种情况下要比遗传算法更有效<sup>[4]</sup>。本文充分利用 PSO 算法具有全局寻优、快速收敛的特点, 将其与 FCM 有机结合, 提出了一种电力系统粒子群优化模糊聚类算法(PSO-FCM)。该算法用 PSO 代替 FCM 中的基于梯度下降的迭代过程, 使算法具有很强的全局搜索能力, 以避免 FCM 易陷

入局部极小的缺陷; 同时也为了降低 FCM 对初始值的敏感度。另外还引入核方法, 将低维特征空间的样本通过核函数映射到高维特征空间, 以增强对特征的优化, 使特征在高维空间更易聚类。文中最后以电力日负荷样本进行了聚类仿真分析。

## 1 基本原理

### 1.1 模糊 C 均值聚类

FCM 是通过求取所有数据点与聚类中心距离加权和目标函数极值点而得到聚类中心迭代公式, 其权值是隶属度函数的指数。

设聚类样本集为:  $X = (x_1, x_2, \dots, x_i, \dots, x_N)$ ,  $x_i$  为任意维向量; 若样本集分为  $K$  个簇, 则簇中心的集合为  $C = (c_1, c_2, \dots, c_K)$ ;  $N \times K$  维矩阵  $U = (u_{ij}), u_{ij} \in [0, 1], i = 1, 2, \dots, N, j = 1, 2, \dots, K$  为样本的隶属度矩阵。因此, FCM 可以表示为目标函数求极值的问题:

$$\min J_m = \sum_{i=1}^N \sum_{j=1}^K u_{ij}^m \|x_i - c_j\|^2 \quad (1)$$

$$\sum_{j=1}^K u_{ij} = 1, u_{ij} \in [0, 1] \quad (2)$$

应用拉格朗日乘法, 结合式(2)的约束条件对式(1)式求导, 得:

$$u_{ij} = \frac{1}{\sum_{k=1}^K \left( \frac{\|x_i - c_j\|^2}{\|x_i - c_k\|^2} \right)^{2/(m-1)}} \quad (3)$$

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m} \quad (4)$$

其中:  $m$  为模糊化程度常数, 是一个控制算法柔性的参数。如果  $m$  过大, 则聚类效果很差; 反之如果  $m$  过小, 则算法接近传统的  $C$  均值算法。

FCM 的基本步骤如下:

1) 根据聚类数目  $K$ , 随机给定簇中心  $C^{(0)} = (c_1, c_2, \dots, c_K)$  以及终止误差  $\varepsilon (\varepsilon > 0)$ , 并设当前迭代次数  $t=0$ ;

2) 按式(3), 由  $C^{(t)}$  求取各样本隶属度, 构成隶属度矩阵  $U^{(t)}$ ;

3) 由式(4)求取下一迭代的簇中心  $C^{(t+1)}$ ;

4) 如果  $\|U^{(t+1)} - U^{(t)}\| < \varepsilon$  则结束迭代, 否则置

$t=t+1$ , 返回步骤 2)。

FCM 算法由于计算簇中心时是基于传统的  $C$  均值算法, 无法避免  $C$  均值算法的一些缺点, 如: 初始簇中心选在一个局部最优点附近, 将会造成算法收敛于局部最小, 并且对噪声数据相当敏感。

### 1.2 核方法

对于许多在低维特征空间线性不可分的样本, 引入核方法, 可将低维特征空间的样本通过核函数映射到高维特征空间, 增加对特征的优化, 使其特征在高维特征空间线性可分, 更易聚类。核函数表示样本从低维特征空间映射到某一高维特征空间  $H$  的非线性映射  $\Phi$  的点积形式:

$$K(x_i, x_j) = (\Phi(x_i), \Phi(x_j)) \quad (5)$$

常用的核函数为高斯核函数, 如:

$$K(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (6)$$

式中:  $\sigma$  为高斯核函数的宽度。该函数对应的特征空间是无穷维的, 有限的样本在其特征空间肯定是线性可分的。

引入核方法, 则式(1)可以转化为:

$$\min J_H = \sum_{i=1}^N \sum_{j=1}^K u_{ij}^m Q_{ij} \quad (7)$$

其中:  $Q_{ij}$  表示在高维特征空间  $H$  中第  $i$  个样本到第  $j$  类簇中心的欧拉距离, 即:

$$Q_{ij} = \|\Phi(x_i) - \Phi(c_j)\|^2 = K(x_i, x_i) + K(c_j, c_j) - 2K(x_i, c_j) \quad (8)$$

在高维特征空间  $H$  中, 新的隶属度函数为:

$$u_{ij} = \frac{1}{\sum_{k=1}^K \left( \frac{Q_{ij}}{Q_{ik}} \right)^{2/(m-1)}} \quad (9)$$

新的聚类中心为:

$$\Phi(c_j) = \sum_{i=1}^N u_{ij}^m \Phi(x_i) / \sum_{i=1}^N u_{ij}^m \quad (10)$$

### 1.3 粒子群优化算法

PSO 算法是一类新兴的基于群智能的随机优化算法。在 PSO 中, 每个粒子都是解空间 ( $n$  维) 的一点, 并且都具有一个速度 ( $n$  维的矢量), 不同粒子具有对应于与目标函数相关的个体适应度。每个粒子根据自身的飞行经验和群体的飞行经验来调整自己的飞行轨迹, 向最优点靠拢。对于某个粒子  $i$ , 它的位置表示为  $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ ; 飞行速度表示为  $v_i = (v_{i1}, v_{i2}, \dots, v_{in})$ ; 它所经历过的最好位置

记为  $p_i = (p_{i1}, p_{i2}, \dots, p_{in})$ ; 它所能获取的群体的最好位置为  $p_g = (p_{g1}, p_{g2}, \dots, p_{gn})$ 。在每一步中, 粒子根据式 (11)、(12) 更新自己的速度和位置。

$$v_{id}(t+1) = wv_{id}(t) + c_1r_1(p_{id}(t) - x_{id}(t)) + c_2r_2(p_{gd}(t) - x_{id}(t)) \quad (11)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1) \quad (12)$$

其中:  $w$  称为惯性权重;  $d=1, 2, \dots, n$ ;  $t$  为当前进化代数;  $i=1, 2, \dots, s$ ,  $s$  为种群规模;  $r_1$  和  $r_2$  为分布于  $[0, 1]$  之间的随机数;  $c_1$  和  $c_2$  为两个正常数, 称为加速因子;  $P_{id}(t)$  为粒子的个体极值位置,  $P_{gd}(t)$  为粒子的全局极值位置。

PSO 中, 惯性权重  $w$  的选择很关键, 较大的  $w$  有较好的全局搜索能力, 而较小的  $w$  有较强的局部搜索能力。因此, 随着迭代次数的增加, 惯性权重应不断减少, 从而使得 PSO 算法在初期具有较强的全局收敛能力, 而晚期具有较强的局部收敛能力。本文  $w$  将随进化代数  $t$  从最大惯性权重  $W_{\max}$  递减至最小惯性权重  $W_{\min}$ , 如式 (13) 所示。

$$w = \left( \frac{T_{\max} - t}{T_{\max}} \right)^{w_n} (W_{\max} - W_{\min}) + W_{\min} \quad (13)$$

式中:  $T_{\max}$  为最大进化代数;  $W_n$  为惯性因子衰减指数; 文中应用时  $W_n$  取 1,  $W_{\max}$  取 0.9,  $W_{\min}$  取 0.3。由于加速因子  $c_1$  较大时, 会使粒子过多地在局部范围徘徊, 而  $c_2$  较大时会促使粒子过早收敛到局部最小值, 为了平衡随机因素的作用, 文中取  $c_1=c_2=1$ 。

## 2 电力系统粒子群优化模糊聚类算法

### 2.1 基本思想

虽然 FCM 算法的每步迭代都沿着好的方向进行, 但这种基于梯度下降的 FCM 算法本质上是一种局部搜索算法, 容易陷入局部极小值, 对于聚类样本数量较大的情况更明显。而且 FCM 算法所能寻找到的最优解对初始值也有很大的依赖性, 即初始分类情况对最终的聚类效果有很大的影响。FCM 算法要想得到好的聚类效果需要对不同的初始值进行多次试验, 从中选取较好的结果, 这既费时, 效率又低。实际上 FCM 算法可归结为目标函数的求极值优化问题。对这种优化问题的处理, PSO 算法具有很强的优势, 因为基于种群操作的 PSO 算法初始为均匀分布空间中的若干可能值, 具有很强全局搜索能力, 不易陷入局部区域, 且收敛速度迅速。因此可以用 PSO 代替 FCM 的迭代过程, 二者结合将能获得性能更加卓越的聚类效果。

### 2.2 粒子的组成

粒子的组成和遗传算法中的染色体组成相似, 粒子的组成表示了粒子在解空间中的位置。FCM 算法的核心是寻找最优聚类中心的问题, 因此本算法把聚类中心  $C_i$  ( $i=1, 2, \dots, K$ ) 作为粒子进行编码。粒子由  $K$  个聚类中心组成, 即  $K$  个聚类中心按顺序排列构成一个粒子。粒子的长度是  $K \times D$ , 其中  $D$  为聚类中心的维数。任一粒子  $x_i$  编码如下:

$$x_i = (c_{i,1}, \dots, c_{i,D}, \dots, c_{i,j}, \dots, c_{K,1}, \dots, c_{K,D})$$

其中:  $c_{i,j}$  表示第  $i$  种聚类方式中的第  $j$  个聚类中心。

### 2.3 适应度函数和终止条件

对于每个粒子的评价, 本文定义如下的适应度函数:

$$f(x_i) = \frac{1}{J_H + 1} \quad (14)$$

如果聚类效果越好, 则  $J_H$  越小, 粒子适应度  $f(x)$  就越高。

由于每个粒子代表的是一种聚类中心的选取, 所以每个粒子适应度的高低表明了在这种聚类中心选取下的聚类效果的优劣。对于算法中每个粒子的评价应该从整个样本空间角度进行, 无需对每个粒子的每个维度进行选取操作评价。

PSO 算法的终止条件为: 满足最大进化代数或  $J_H$  小于规定的  $\varepsilon$  或模糊矩阵不变。模糊矩阵不变可定义为:

$$|u_{ij}^t - u_{ij}^{t-1}| \leq \varepsilon \quad (15)$$

### 2.4 算法实现过程

1) 初始化参数。首先, 给定模糊化程度常数  $m$ , 聚类簇数  $K$ , 以及粒子群群体规模  $s$ 。随机选出  $K$  个样本作为一个聚类中心集 (即粒子), 共进行  $s$  次, 产生  $s$  个粒子, 并随机产生每个粒子的更新速度  $v_i$ 。

2) 评价种群  $X(t)$ 。由适应度函数式 (14), 计算每个粒子的适应度, 再由式 (16) 求取个体极值位置  $p_{id}(t)$ ; 如果在当前种群的所有粒子中, 存在粒子的个体极值位置  $p_{id}(t)$  的适应度大于全局极值位置  $p_{gd}(t)$  的适应度, 则更新  $p_{gd}(t)$  为该粒子的  $p_{id}(t)$ , 否则  $p_{gd}(t)$  不变。

$$p_{id}(t) = \begin{cases} x_{id}(t), & f(x_{id}(t)) \geq f(p_{id}(t-1)) \\ p_{id}(t-1), & f(x_{id}(t)) < f(p_{id}(t-1)) \end{cases} \quad (16)$$

3) 检查终止条件, 若满足, 则计算结束, 转至步骤 4); 否则, 由式 (11) 和式 (12) 计算新一代粒子的

速度  $v_i(t+1)$  和位置  $x_i(t+1)$ , 再转至步骤 2)。

4) 输出终止位置的粒子, 即聚类中心的集合。

### 3 算法应用分析

为了验证 PSO-FCM 算法的有效性, 本文将其应用于负荷预测样本数据的预处理。基于神经网络、支持向量机等负荷预测算法的预测精度主要依赖于训练样本集的选择, 合理的输入特征向量可使预测算法快速、有效逼近目标矢量, 达到预测误差要求, 使预测算法具有良好的泛化性能。

目前, 许多文献已应用了一些辅助策略来改善负荷预测精度。如采用相似日法<sup>[5]</sup>, 但相似日法由于负荷数据的日期类型难于确定, 如长假期及节假日日期不定, 将造成负荷模式难以统一, 在实际应用中还需要调度人员借助经验进行调整。

本文以 EUNITE 历史负荷为样本<sup>[6]</sup>, 分别使用 FCM 和 PSO-FCM 进行聚类分析仿真实验。采用 1997 年 1 月 1 日至 1998 年 12 月 31 日共两年 730 个日负荷数据样本进行日负荷曲线聚类, 日负荷序列相量维数为 48 维。

在仿真实验中, 模糊化程度常数  $m$  取 2, 粒子群规模取 20, 惯性权值  $w$  的初始值取 0.9, 最大进化代数  $T_{max}$  取 150, 高斯核函数的宽度  $\sigma$  取 3。

聚类分析时, 必须先确定聚类中心数  $K$  的大小, 随意确定  $K$  的大小会导致聚类效果的劣化。本文采取文献[7]提出的聚类有效性函数 (即聚类的紧致性函数  $comp$  和分离性函数  $sep$  之比:  $S(U,K)=comp/sep$ ) 来确定训练样本的聚类个数, 在此  $K$  取值为 10。

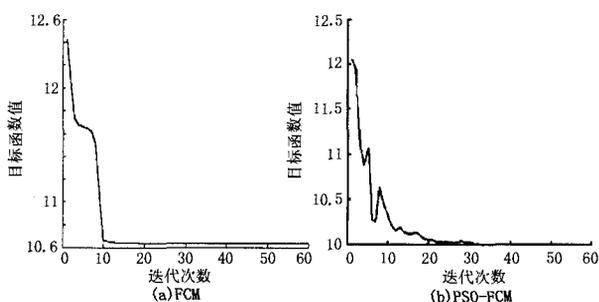


图 1 目标函数值随迭代次数变化情况

Fig.1 Change of object function value with iterative times

对两种聚类算法分别作 30 次仿真实验。图 1(a) 和图 1(b) 分别为某次 FCM 和 PSO-FCM 聚类仿真过程目标函数随迭代次数变化情况。在样本的聚类过程中, 对于 FCM 算法, 每次聚类过程得到的目标函数值随迭代次数变化的情况基本上都与图 1(a) 一致, 可以看出 FCM 的目标函数值在逐步减小的过

程中没有出现反复情形, 这也说明基于梯度下降的 FCM 算法在样本空间进行聚类时有可能陷入局部极小值。而对于 PSO-FCM 算法在聚类过程中目标函数值是随迭代次数变化的, 并不是梯度下降, 因此不易陷入局部极小值。

对于 FCM, 给定不同的初始值, 将收敛于不同的目标函数极值, 且差异很大, 因此 FCM 对初始值敏感, 易陷入局部最优。而对于 PSO-FCM, 给与不同的初始值, 虽然每次迭代次数和目标函数的变化过程不一样, 但是都将趋于稳定地收敛于相同的目标函数极值。30 次仿真实验的目标函数极值情况如图 2 所示。由图 2 可见, 对于不同实验, FCM 算法的目标函数极值变化很大, 该算法是不稳定的; 而 PSO-FCM 算法目标函数极值则是稳定的。

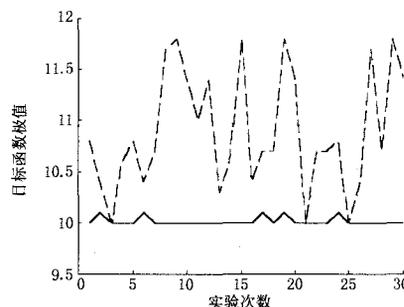


图 2 目标函数极值情况

Fig.2 Object function convergence

本文采用分离系数、分离熵以及紧致性与分离性效果三项指标评价 FCM 和 PSO-FCM 聚类算法聚类效果。

分离系数定义为:

$$F(U, K) = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K (u_{ij})^2 \quad (17)$$

分离系数的含义: 表征所有输入样本相对于聚类中心的接近程度。如果每个样本仅属于一类, 此聚类效果最佳, 此时  $F(U,K)=1$ ; 否则  $F(U,K)<1$ 。因此, 分离系数越大聚类效果越好。

分离熵定义为:

$$H(U, K) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K u_{ij} \log(u_{ij}) \quad (18)$$

若所有的  $u_{ij}$  接近 0 或 1, 则熵就小, 所得的聚类效果就好; 若  $u_{ij}$  接近  $1/K$ , 则聚类块的模糊程度高, 熵就大, 聚类效果就差。

紧致性与分离性效果(有效性函数)定义为:

$$S(U, K) = \frac{\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^K u_{ij}^2 \|x_i - c_j\|^2}{\min_{i,j} \|c_i - c_j\|^2} \quad (19)$$

$S(U, K)$ 是输入样本与它们相应的聚类中心间距的平均值与聚类中心最小间距的比值。一个好的聚类应当使聚类中心的间距尽可能地大,而样本与其中心的间距尽可能地小。因此  $S(U, K)$  越小, 聚类效果越好。

根据 30 次聚类仿真的结果计算三种评价准则的平均值, 如表 1 所示。可见: 从三项指标均表明 PSO-FCM 算法聚类效果优于 FCM 算法。

表 1 评价结果

Tab.1 Evaluation result

算法	分离系数	分离熵	有效性函数
FCM	0.653	11.401	42.367
PSO-FCM	0.701	10.893	26.985

限于篇幅, 在此仅列出 PSO-FCM 聚类结果的隶属各类的样本个数, 如表 2 所示。图 3 给出了 PSO-FCM 算法聚类获得的聚类中心日负荷曲线。

表 2 PSO-FCM 聚类结果

Tab.2 Cluster results by PSO-FCM

序号	个数	序号	个数
1	67	6	62
2	62	7	52
3	86	8	101
4	59	9	105
5	73	10	63

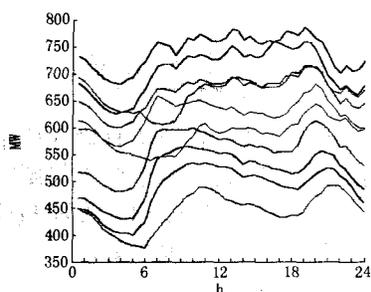


图 3 聚类中心曲线

Fig.3 Curve of cluster center

## 4 结论

本文提出的电力系统粒子群优化模糊聚类算法(PSO-FCM), 充分利用了 PSO 具有的强全局搜索能力; 同时还通过核方法将低维特征空间的样本通过核函数映射到高维特征空间, 增强了特征的优化, 使特征在高维空间更易聚类。电力系统负荷样

本聚类的应用仿真结果表明: 该方法能有效地避免 FCM 易陷入局部极小的缺陷, 同时也降低了 FCM 对初始值的敏感度, 从而提高了聚类的效率和效果。

## 参考文献

- [1] 马瑞, 贺仁睦. 一种基于粒子群优化并行神经网络的电力系统负荷特性聚类方法[J]. 现代电力, 2006, 23(3): 1-5.  
MA Rui, HE Ren-mu. A Clustering Approach of Power System Load Based on Parallel Neural Network with Particle Swarm Optimization[J]. Modern Electric Power, 2006, 23(3): 1-5
- [2] 孙雅明, 王晨力, 张智晟, 等. 基于蚁群优化算法的电力系统负荷序列的聚类分析[J]. 中国电机工程学报, 2005, 25(18): 40-45.  
SUN Ya-ming, WANG Chen-li, ZHANG Zhi-sheng, et al. Clustering Analysis of Power System Load Series based on Ant Colony Optimization Algorithm[J]. Proceedings of the CSEE, 2005, 25(18): 40-45.
- [3] 李培强, 李欣然, 陈辉华, 等. 基于模糊聚类的电力负荷特性的分类与综合[J]. 中国电机工程学报, 2005, 25(24): 73-78.  
LI Pei-qiang, LI Xin-ran, CHEN Hui-hua, et al. The Characteristics Classification and Synthesis of Power Load Based on Fuzzy Method [J]. Proceedings of the CSEE, 2005, 25(24): 73-78.
- [4] Kennedy J. The Particle Swarm: Social Adaptation of Knowledge[A]. In: Proc IEEE Int Conf Evol Comput[C]. Indianapolis: 1997. 303-308.
- [5] 莫维仁, 张伯明, 孙宏斌, 等. 短期负荷预测中选择相似日的探讨[J]. 清华大学学报(自然科学版), 2004, 44(1): 106-109.  
MO Wei-ren, ZHANG Bo-ming, SUN Hong-bin. Method to Select Similar Days for Short-term Load Forecasting [J]. Journal of Tsinghua University (Sci&Tech), 2004, 44(1): 106-109.
- [6] World Wide Competition within the EUNITE Network, EUNITE Competition Report[EB/OL]. [Http://neuron.tuke.sk/competition/index.php](http://neuron.tuke.sk/competition/index.php).
- [7] Xie X L, Beni G. A Validity Measure for Fuzzy Clustering[J]. IEEE Trans on Patt Anal, Machine Intell, 1991, 13(8): 1-5.

收稿日期: 2007-04-29

作者简介:

何晓峰(1979-), 男, 博士研究生, 研究方向为电力系统调度自动化和数据仓库;

王 钢(1966-), 男, 博士, 教授, 博士生导师, 主要从事电力系统保护、控制及自动化领域的教学与科研工作; E-mail: wg\_scut@263.net

李海锋(1976-)男, 博士, 讲师, 主要从事电力系统故障分析与继电保护的教学与科研工作。