

基于离散 Fréchet 距离和 LS-SVM 的短期负荷预测

陈超¹, 黄国勇^{1,2}, 范玉刚^{1,2}, 吴建德^{1,2}, 王晓东^{1,2}

(1. 昆明理工大学信息工程与自动化学院, 云南 昆明 650500; 2. 云南省矿物管道输送工程技术研究中心, 云南 昆明 650500)

摘要: 针对现有电力系统短期负荷预测精度低、数据处理量大、易受输入变量的影响等问题, 提出了一种将离散 Fréchet 距离与 LS-SVM 相结合的短期负荷预测方法。分析总结了 East-Slovakia Power Distribution Company 提供的历年负荷数据, 结合该地区的用电规律, 通过引入离散 Fréchet 距离, 建立离散曲线相似性的数学模型, 选取与基准曲线形状相似的相似日, 利用相似日负荷数据对 LS-SVM 预测模型进行训练。经过仿真验证, 并与标准 LS-SVM 模型得到的结果对比, 所提预测方法明显提高了预测精度。

关键词: 离散 Fréchet 距离; LS-SVM; 用电规律; 形状相似日; 短期负荷预测

Short-term load forecasting based on discrete Fréchet distance and LS-SVM

CHEN Chao¹, HUANG Guo-yong^{1,2}, FAN Yu-gang^{1,2}, WU Jian-de^{1,2}, WANG Xiao-dong^{1,2}

(1. School of Information Engineering and Automation, Kunming University of Technology, Kunming 650500, China;

2. Yunnan Mineral Pipeline Transmission Engineering Technology Research Center, Kunming 650500, China)

Abstract: In order to improve the precision of short-term load forecasting and overcome the disadvantage of large amount of data and the influence of input variables, a new method based on the combination of discrete Fréchet distance and LS-SVM is presented. This paper analyzes and summarizes the historical load data provided by the East-Slovakia Power Distribution Company. Combining with the law of the region's electricity, by introducing discrete Fréchet distance, the shape-similar days which are similar to the reference day are selected by establishing the mathematical model of discrete curve similarity, and then the similar daily load data are used to train the LS-SVM forecasting model. Through simulation and comparison with the results obtained by the standard LS-SVM model, it is proved that the prediction methods significantly improve the prediction accuracy.

This work is supported by National Natural Science Foundation of China (No. 51169007).

Key words: discrete Fréchet distance; LS-SVM; electricity regularity; shape similar days; short-term load forecasting

中图分类号: TM714 文献标识码: A 文章编号: 1674-3415(2014)05-0142-06

0 引言

电力系统的短期负荷预测是负荷预测的重要组成部分, 它主要用于预报未来几小时、1 天至几天的电力负荷, 对于调度部门的机组最优组合、经济调度而言, 尤其是对现在与将来的电力市场有着重要的意义^[1]。近十几年来, 众多研究人员提出了很多有效的预测方法, 如灰色模型预测法^[2-3]、人工神经网络^[4]、支持向量机^[5]等。

由于短期负荷预测所具有的特点, 相似日方法

成为了短期负荷预测的基本方法之一。目前, 相似日的选取一般从两个方面考虑: 各种影响因素分析和负荷曲线的相似度。文献[6]通过模糊聚类技术将不同用户的负荷特性曲线进行分类, 利用与预测曲线相同类型的典型曲线结合相关的影响因素, 建立了 BP 网络预测模型。文献[7]深入分析了短期负荷预测中的主要影响因素, 提出基于相似日负荷修正的算法, 并运用到节假日的短期负荷预测上。文献[8]深入分析了常见因素对负荷影响的一般规律, 给出了各种影响因素的相似度函数, 再对这些相似度进行组合, 以求取相似日。文献[9]通过标准化影响负荷的各因素, 建立系数特征映射表, 利用规范化后的特征数据选择相似日。文献[10]用趋势相似度和形状相似度选择相似日, 选取了合适的相似日。

基金项目: 国家自然科学基金资助项目 (51169007); 云南省科技计划项目 (2011DA005 & 2011FZ036 & 2012CA022); 云南省中青年学术和技术带头人后备人才培养计划项目 (2011CI017)

文献[11]运用混沌相空间重构法提取负荷的平衡点、准周期, 找到由当前特征聚类出的各个历史相似日。文献[10-11]在选取相似日时均考虑了日负荷的平均值和日负荷曲线形状。

基于以上分析, 本文对 East-Slovakia Power Distribution Company 提供的电网数据进行分析总结, 得出了该地区的用电特点, 通过引入离散 Fréchet 距离判别曲线相似性的方法来选取负荷曲线形状相似日, 利用相似日的负荷数据对 LS-SVM 进行训练, 得到预测模型。仿真验证表明, 该方法能够从负荷曲线形状相似性的角度较好地选取出相似日, 且预测的精度较高。

1 离散 Fréchet 距离和 LS-SVM 理论

1.1 离散 Fréchet 距离

Fréchet 距离是一种判别曲线之间相似性的距离测度, 文献[12-14]都曾将 Fréchet 距离应用到判别两条曲线的相似性上, 并且取得了较好的效果。而离散 Fréchet 距离是 Eiter 和 Mannila 在连续 Fréchet 距离的基础上提出的, 本文将离散 Fréchet 距离作为负荷曲线中关键特征至高点与至低点的距离测度来研究曲线的相似性。其定义如下:

定义 1^[15]: (1) 给定一个有 n 个至高点的多边形链 $P = \{P_1, P_2, \dots, P_n\}$, 一个沿着 P 的 k 步, 分割 P 的至高点成为 k 个不相交的非空子 $\{P_i\}_{i=1, \dots, k}$, 使得 $P_i = \langle P_{n_{i+1}}, \dots, P_{n_i} \rangle$ 和 $1 = n_0 < n_1 < \dots < n_k = n$ 。

(2) 给定两个多边形链 $A = \langle a_1, \dots, a_m \rangle$, $B = \langle b_1, \dots, b_n \rangle$, 一个沿着 A 和 B 的组合步是一个沿着 A 的 k 步 $\{A_i\}_{i=1, \dots, k}$ 和一个沿着 B 的 K 步 $\{B_i\}_{i=1, \dots, k}$ 组成, 使得对于 $1 \leq i \leq k$, 且 A_i, B_i 中有一个恰好包含一个至高点。

(3) 一个沿着链 A 和 B 的组合步 $W = \{(A_i, B_i)\}$ 的花费(cost)就是

$$d_F^W(A, B) = \max_i \max_{(a, b) \in A_i \times B_i} \text{dist}(a, b) \quad (1)$$

则链 A 和 B 间的离散 Fréchet 距离就是

$$d_F(A, B) = \min_W d_F^W(A, B) \quad (2)$$

这个组合步称为链 A 和 B 的 Fréchet 排列。Fréchet 排列表现为两条曲线上离散点的对应情况, 如果两条曲线上的离散点是固定的, 那么可以有很多种不同的 Fréchet 排列, 且每种 Fréchet 排列都对应一个相应的离散 Fréchet 距离, 然后再根据实际应用选择出最好的 Fréchet 排列以及对应的离散 Fréchet 距离。

但是, 离散 Fréchet 距离只是表示出两条曲线关键特征点间的距离, 还不能够真正判别出曲线的相似度, 因此, 本文采用定义 2 来判别曲线的相似度。

定义 2^[16]: 设 $A = \langle a_1, \dots, a_m \rangle$, $B = \langle b_1, \dots, b_n \rangle$ 组成的曲线, 称 $d_F^1(A, B)$ 为它们的至高点间的离散 Fréchet 距离, $d_F^2(A, B)$ 为它们的至低点间的离散 Fréchet 距离; 如果对于给定的值 ε , 如果 A, B 相似, 则 $|d_F^1(A, B) - d_F^2(A, B)| \leq \varepsilon$, 否则称它们不相似。

1.2 LS-SVM 理论

LS-SVM 是对标准 SVM 的一种扩展, 它将 SVM 中的不等式约束转换成等式约束, 使得 SVM 的求解由解决一个二次规划问题转变为一个线性方程组的求解问题, 极大地简化了计算, 提高了模型的训练效率。

在 LS-SVM 算法中, 输入空间的函数估计可转换为如下的最优问题。

$$\begin{aligned} \min \phi &= \frac{1}{2} \|w\|^2 + \frac{1}{2} C \sum_{i=1}^l \xi_i^2 \\ \text{s.t. } & y_i = w \cdot \varphi(x_i) + b + \xi_i \end{aligned} \quad (3)$$

式中: w 为权值向量; ξ_i 为松弛变量; C 为惩罚因子; b 为偏差值。在引入 Lagrange 函数后, 根据 Mercer 条件和 KKT 定理可以得到矩阵方程。

$$\begin{bmatrix} \mathbf{0} & \mathbf{1}^T \\ \mathbf{1} & \Omega + C^{-1} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \end{bmatrix} \quad (4)$$

其中: $\mathbf{y} = [y_1, \dots, y_l]$, $\alpha = [\alpha_1, \dots, \alpha_l]$, Ω 中的元素为 $\Omega_{ij} = K(x_i, x_j)_{i, j=1, 2, \dots, l}$ 。求解式 (4) 后得到 LS-SVM 的函数估计为:

$$f(x) = \sum_{i=1}^l \alpha_i K(x_i, x_j) + b \quad (5)$$

本文中, 选取的核函数 $K()$ 为径向基函数。

2 用电规律分析

本文根据 East-Slovakia Power Distribution Company 提供的历年电网数据得到该地区每个月的日平均负荷曲线, 通过统计分析发现, 由于该地区季节气温变化明显, 负荷趋势受气象因素显著, 总的来说, 每年 2、3 月份和 4 月上旬的负荷水平趋势相近, 从 4 月下旬开始负荷逐渐变小, 5~9 月份的负荷相对较低且水平趋势相近, 从 10 月开始负荷

逐渐变大, 11~1 月份的负荷水平趋势又相近, 而且 12 月份和 1 月份的负荷能够达到全年最高。图 1 所示为 5~9 月份与 11~1 月份的典型日平均负荷曲线的对比, 从图中可看出, 11~1 月份的平均负荷明显要高于 5~9 月份。因此, 本文根据这一规律和预测日所属月份, 在对应时间段内进一步选取相似日, 这样既缩小了样本选取范围, 也使得相似日的选取更具有针对性。

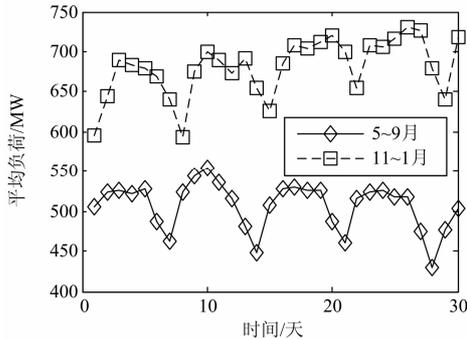


图 1 日平均负荷曲线对比

Fig. 1 Comparison of average daily load curves

其次, 对每日 24 点的负荷曲线分析还发现, 在判断形状相似性时, 如果两天的负荷曲线是相似的话, 它们的至高点(至低点)的数量相差不超过 2 个, 且其中一条的某个峰值点只可能与另一条曲线上同一时刻或相差 1~2 h 的峰值点对应, 这也体现了人们生活习惯的规律性。利用这一特性, 可大大减少 Fréchet 排列的情况, 同时加快了离散 Fréchet 距离的计算速度。图 2 所示为两个形状相似日的曲线对比, 从图中可看出, 两天的负荷曲线的峰值点数量相同, 且几乎出现在同一时刻。

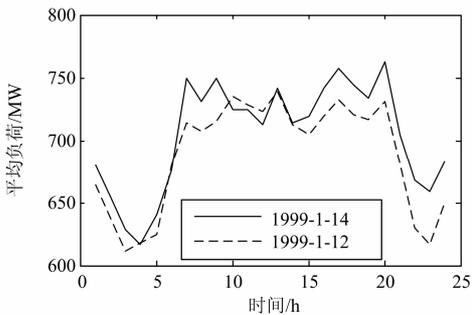


图 2 形状相似曲线对比

Fig. 2 Comparison of shape-similar curves

3 预测方法

3.1 相似日的选取

本文将样本数据以 7 天为周期进行划分, 把预测日前 6 天的负荷曲线作为基准曲线进行相似日选

取, 例如: 预测日为星期二, 基准曲线就为上星期三到星期一的负荷曲线, 然后将历史日中每条星期三到星期一的负荷曲线与基准曲线作相似度计算。

在选取相似日时, 选取负荷曲线的至高点 and 至低点作为关键特征, 其定义的标准是: 一个点如果它比前两个点大(小)同时还比后两个点大(小), 就称它为一个至高点(至低点)。具体的相似日选取步骤如下:

(1) 将预测日前 6 天作为基准日, 分别找出基准曲线和历史日曲线的至高点与至低点, 将它们表示成 $A = \langle a_1, \dots, a_m \rangle$, $B = \langle b_1, \dots, b_n \rangle$ (a_1, \dots, a_m 是曲线 A 的 m 个至高点(至低点), b_1, \dots, b_n 是曲线 B 的 n 个至高点(至低点)), 且 $m \leq n$, 若 $n - m > 2$, 则认为它们不相似, 否则执行下一步。

(2) 以峰值点少的曲线 A 为基准, 将峰值点多的曲线 B 划分成 m 步, 划分时尽量保证同一时刻的峰值相互对应, 即 a_i 只能对应空或者与 b_{i-1} , b_i , b_{i+1} 之中的某个峰值点对应, 由于上一步的划分直接影响到下一步的划分, 因此假设有 K 种划分, 每一种划分 $W_j = \{(A_i, B_i)\}, (1 \leq i \leq m, 1 \leq j \leq k)$ 。

(3) 在每一种划分中, 先计算出每一步中所有对应点之间的最大距离, 再求出这种划分的所有步中的最大距离中的最大值, 如式(1)。

(4) 找出所有划分方法中的距离的最小值, 则为曲线 A 和 B 之间的离散 Fréchet 距离, 如式(2)。

(5) 分别得到至高点 and 至低点的最小离散 Fréchet 距离, 让两个最小距离相减, 如果结果的绝对值小于一个特定的阈值 ε , 则两条曲线相似。

3.2 预测模型的建立

假设某一相似时间段内 7 天的负荷为

$$y = \begin{bmatrix} y_{1,1} & y_{2,1} & \dots & y_{7,1} \\ y_{1,2} & y_{2,2} & \dots & y_{7,2} \\ \vdots & \vdots & \vdots & \vdots \\ y_{1,24} & y_{2,24} & \dots & y_{7,24} \end{bmatrix} \quad (6)$$

其中, 列代表一天 24 个点的负荷值, 行表示 7 天同一时刻的负荷值。在训练模型时, 将每个相似时期前 6 天的负荷数据作为输入, 第 7 天(星期类型与预测日相同)的负荷数据作为输出。

为了对预测效果进行合理的评价, 本文选用平均绝对误差和最大误差作为依据。

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - Y'_i}{Y_i} \right| \times 100\% \quad (7)$$

$$ME = MAX |Y_i - Y'_i| \quad (8)$$

式中: Y_i 和 Y'_i 分别为第 i 天的真实值和预测值; n 为负荷采集点数, $n=24$ 。

4 仿真结果与分析

本文采用 East-Slovakia Power Distribution Company 提供的 1998 年每天 24 点的负荷数据进行仿真验证, 并与标准 LS-SVM 的预测结果作对比分析。预测日为 9 月 30 日(星期三), 基准日为 9 月 24 日(星期四)—9 月 29 日(星期二), 根据分析得到的该地区的用电规律, 利用离散 Fréchet 距离在 4 月~9 月的样本中选取相似日, 通过统计分析把阈值 ϵ 设置为 14, 经过计算后, 选取出 13 个相似时段, 如表 1。

表 1 选取出的相似日

Table 1 The selected similar days

相似日	峰值点距离差	相似日	峰值点距离差
9-17--9-23	2	5-28--6-03	8
4-09--4-15	4	5-14--5-20	9
4-23--4-29	4	4-16--4-22	11
4-30--5-06	4	8-27--9-02	11
9-10--9-16	7	7-09--7-15	14
4-02--4-08	8	9-03--9-09	14
5-21--5-27	8		

基准曲线与相似曲线的比较如图 3, 从图 3 可看出, 利用离散 Fréchet 距离选取出的相似曲线, 无论在水平趋势方面还是在曲线形状方面都十分相似。限于篇幅, 本文仅给出比较具有代表性的一条相似曲线, 其余相似曲线同样具有较强的相似性。

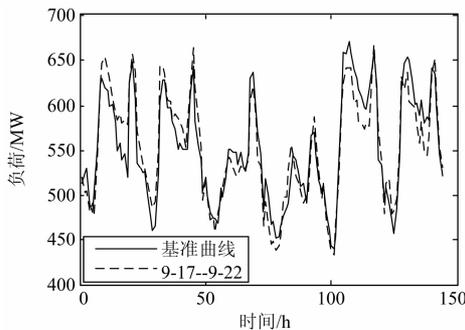


图 3 基准曲线与相似曲线的对比

Fig. 3 Comparison between datum curve and similar curve

在选取出相似日之后, 将各相似日的负荷数据作为训练样本, 对 LS-SVM 模型进行训练。预测结果如图 4, 从图中可看出, 本文的预测方法具有较

高的拟合精度。

为了突出本文方法的优越性, 选用标准 LS-SVM 方法同样对 9 月 30 日的负荷进行预测, 遵循“近大远小”的原则, 选取预测日前 3 个月的负荷数据作为训练样本, 并比较预测精度, 如表 2。

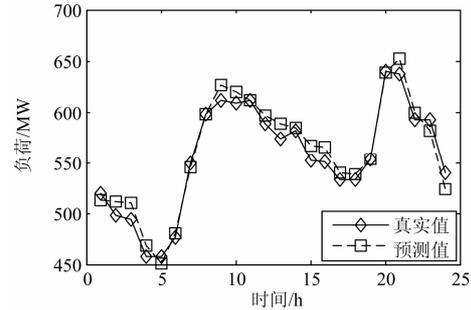


图 4 预测结果

Fig. 4 Prediction results

表 2 预测误差比较

Table 2 Comparison of prediction error

预测时刻	绝对误差		预测时刻	绝对误差	
	本文方法	LS-SVM		本文方法	LS-SVM
00:00	1.156 0	1.087 9	12:00	2.439 0	3.447 8
01:00	2.610 4	7.626 4	13:00	0.515 4	0.138 1
02:00	3.238 8	5.170 3	14:00	2.717 3	1.079 1
03:00	2.407 0	5.190 2	15:00	2.722 3	2.480 0
04:00	1.528 3	3.425 1	16:00	1.313 3	3.525 0
05:00	0.840 3	2.325 1	17:00	1.125 7	3.927 1
06:00	0.728 5	0.298 2	18:00	0.180 8	2.661 8
07:00	0	1.629 9	19:00	0.156 2	0.704 1
08:00	2.618 6	0.155 3	20:00	2.354 7	3.180 4
09:00	1.642 0	0.773 5	21:00	1.011 8	2.154 1
10:00	0	2.329 9	22:00	1.854 9	6.434 1
11:00	1.530 6	0.180 7	23:00	2.962 9	0.460 3

从表 2 中可知, 本文方法的平均绝对误差为 1.5%, 标准 LS-SVM 的平均绝对误差为 2.5%, 本文方法预测值中的 15 个点准确性高于 LS-SVM 预测方法。其次, 本文方法预测的最大误差为 16, LS-SVM 的最大误差为 38, 本文方法明显优于 LS-SVM 预测方法。

5 结论

(1) 本文分析总结了 East-Slovakia Power Distribution Company 提供的历年电网数据, 得到了该地区的负荷规律。结合负荷规律, 引入离散 Fréchet 距离判别曲线相似性的方法, 选取出形状相

似、趋势相似的相似日，减少了训练样本。

(2) 选取出相似日后，用具有更好泛化性的 LS-SVM 进行预测建模，较好地解决了小样本、高维数的问题。

(3) 与标准 LS-SVM 方法得到的预测结果对比分析，本文方法得到的预测精度更高，验证了本文方法的有效性。

参考文献

- [1] 廖旋焕, 胡智宏, 马莹莹, 等. 电力系统短期负荷预测方法综述[J]. 电力系统保护与控制, 2011, 39(1): 147-152.
- LIAO Ni-huan, HU Zhi-hong, MA Ying-ying, et al. Review of the short-term load forecasting methods of electric power system[J]. Power System Protection and Control, 2011, 39(1): 147-152.
- [2] 张志明, 金敏. 基于灰关联分段优选组合模型的短期电力负荷预测研究[J]. 电工技术学报, 2009, 24(6): 115-120.
- ZHANG Zhi-ming, JIN Min. Research on short-term electrical load forecasting based on optimized combination model of grey correlation segmentation[J]. Transactions of China Electrotechnical Society, 2009, 24(6): 115-120.
- [3] 魏少岩, 吴俊勇. 基于灰色模型和 Kalman 平滑器的多母线短期负荷预测[J]. 电工技术学报, 2010, 25(2): 158-162.
- WEI Shao-yan, WU Jun-yong. Short-term bus load forecasting based on grey model and Kalman smoother[J]. Transactions of China Electrotechnical Society, 2010, 25(2): 158-162.
- [4] 祝燕萍, 方鸽飞. 基于动态自适应神经网络和人体舒适度的短期负荷预测[J]. 电力系统保护与控制, 2012, 40(1): 56-61.
- ZHU Yan-ping, FANG Ge-fei. Short-term load forecasting based on dynamic adaptive artificial neural network and human body amenity indicator[J]. Power System Protection and Control, 2012, 40(1): 56-61.
- [5] 黄元生, 邓佳佳. 基于 ARMA 误差修正和自适应粒子群优化的 SVM 短期负荷预测[J]. 电力系统保护与控制, 2011, 39(14): 26-32.
- HUANG Yuan-sheng, DENG Jia-jia. SVM short-term load forecasting based on ARMA error calibration and adaptive particle swarm optimization[J]. Power System Protection and Control, 2011, 39(14): 26-32.
- [6] 黎祚, 周步祥, 林楠. 基于模糊聚类与改进 BP 算法的日负荷特性曲线分类与短期负荷预测[J]. 电力系统保护与控制, 2012, 40(3): 56-60.
- LI Zuo, ZHOU Bu-xiang, LIN Nan. Classification of daily load characteristics curve and forecasting of short-term load based on fuzzy clustering and improved BP algorithm[J]. Power System Protection and Control, 2012, 40(3): 56-60.
- [7] 林辉, 刘晶, 郝志峰, 等. 基于相似日负荷修正的节假日短期负荷预测[J]. 电力系统保护与控制, 2010, 38(7): 47-51.
- LIN Hui, LIU Jing, HAO Zhi-feng, et al. Short-term load forecasting for holidays based on the similar day's load modification[J]. Power System Protection and Control, 2010, 38(7): 47-51.
- [8] 黎灿兵, 李晓辉, 赵瑞, 等. 电力短期负荷预测相似日选取算法[J]. 电力系统自动化, 2008, 32(9): 69-73.
- LI Can-bing, LI Xiao-hui, ZHAO Rui, et al. A novel algorithm of selecting similar days for short-term load forecasting[J]. Automation of Electric Power Systems, 2008, 32(9): 69-73.
- [9] 刘梦良, 刘晓华, 高荣. 基于相似日小波支持向量机的短期电力负荷预测[J]. 电工技术学报, 2006, 21(11): 59-64.
- LIU Meng-liang, LIU Xiao-hua, GAO Rong. Short-term load forecasting using wavelet transform and SVM based on similar-days[J]. Transactions of China Electrotechnical Society, 2006, 21(11): 59-64.
- [10] 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, et al. Method to select similar days for short-term load

- forecasting[J]. J Tsinghua Univ: Sci & Tech, 2004, 44(1): 106-109.
- [11] 杨正瓴, 田勇, 张广涛, 等. 相似日短期负荷预测的非线性理论基础与改进[J]. 电网技术, 2006, 30(6): 63-66. YANG Zheng-ling, TIAN Yong, ZHANG Guang-tao, et al. Nonlinear theoretical foundation and improvement of similar days method for short term load forecasting[J]. Power System Technology, 2006, 30(6): 63-66.
- [12] ALT H, GODAN M. Computing the Fréchet distance between two ploygonal curves[J]. International Journal of Computational Geometry & Applications, 1995, 5: 75-91.
- [13] ALT H, KNAUER C, WENK C. Matching polygonal curves with respect to the Fréchet distance[C] // Proceedings of the 18th Annual Symposium on Theoretical Aspects of Computer Science (STACS'01) (LNCS2010). London: Springer-Verlag, 2001: 63-74.
- [14] MOSIG A, CLAUSEN M. Approximately matching polygonal curves with respect to the Fréchet distance[J]. Computational Geometry: Theory and Applications, 2005, 30(2): 113-127.
- [15] EITER T, MANNILA H. Computing discrete Fréchet distance[R]. Vienna: Technical University of Vienna, 1994.
- [16] 朱洁, 黄樟灿, 彭晓琳. 基于离散 Fréchet 距离的判别曲线相似性的算法[J]. 武汉大学学报: 理学版, 2009, 55(2): 227-232. ZHU Jie, HUANG Zhang-can, PENG Xiao-lin. Curve similarity judgment based on the discrete Fréchet distance[J]. J Wuhan Univ: Nat Sci Ed, 2009, 55(2): 227-232.

收稿日期: 2013-06-04

作者简介:

陈超(1989-), 男, 硕士研究生, 从事电力系统分析与控制研究; E-mail:zxycc611@163.com

黄国勇(1977-), 男, 博士, 副教授, 从事智能信息处理、智能控制研究。