

基于混沌时间序列局域法的短时交通流预测

廖荣华^{1,2}, 兰时勇^{1,2}, 刘正熙^{1,2}

(1. 四川大学 计算机学院, 四川 成都 610065;

2. 四川大学 视觉合成重点学科实验室, 四川 成都 610045)

摘要:为提高城市短时交通流预测精度,将混沌时间序列分析应用于城市短时交通流数据,研究混沌时间序列局域预测法中的加权零阶局域法和加权一阶局域法。针对局域预测法在选取邻近相点时采用的欧氏距离和向量夹角两种方法只能片面反映邻近点的特点的问题,提出一种改进邻近相点选取的方法,综合相点相似程度和相点距离来选取邻近相点。再将原有方法和改进后的方法应用于北京市短时交通流预测中。结果表明,混沌时间序列局域法能适用于短时交通流预测,并且改进后的方法比原有方法具有更高的预测精度。

关键词:交通流预测;混沌时间序列;邻近点;加权零阶局域法;加权一阶局域法

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2015)01-0001-05

doi:10.3969/j.issn.1673-629X.2015.01.001

Short-term Traffic Flow Forecasting Based on Local Prediction Method in Chaotic Time Series

LIAO Rong-hua^{1,2}, LAN Shi-yong^{1,2}, LIU Zheng-xi^{1,2}

(1. College of Computer, Sichuan University, Chengdu 610065, China;

2. State Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu 610045, China)

Abstract: To improve the accuracy of urban short-term traffic flow forecasting, the chaotic time series analysis is applied to urban short-term traffic flow data, study the two local chaotic time series prediction, including adding-weight zero-rank local-region method and adding-weight one-rank local-region method. Euclidean distance method and vector angle method used in selecting neighbor points in local prediction method are being researched, and these two methods can not reflect the overall characteristics of the neighbor points, in view of this problem, an improved neighboring phase point selection method which integrated relative degree of similarity and distance to select neighbor phase points is presented. Then the old methods and the improved method are used in the Beijing short-term traffic flow prediction. The results show that local prediction method in chaotic time series can be used in short-term traffic flow forecasting, and the improved method has higher accuracy in prediction than the old methods.

Key words: traffic flow forecasting; chaotic time series; neighbor point; adding-weight zero-rank local-region method; adding-weight one-rank local-region method

0 引言

实现智能交通系统控制和诱导的关键是实时、准确的交通流预测。交通流预测中,一般都是用短时交通流预测。短时交通流预测一般是指观测时间不超过15 min的交通流预测。短时交通流预测的方法有很多,文献[1-4]提出基于数理统计的方法:时间序列预测法、卡尔曼滤波预测法、历史平均预测法等;文献[5

-8]提出基于智能模型的预测方法:非参数回归分析预测法、支持向量机回归预测法、神经网络预测法等。由于交通系统本身是一个不确定的系统,各种因素交织在一起使得系统变得非常复杂,而传统数理统计方法只研究交通流的线性特性,忽略其中的不确定性;智能模型的方法计算复杂且计算量大,因此上述方法都有其局限性。文献[9-13]将混沌理论引入交通流研

收稿日期:2014-02-12

修回日期:2014-05-15

网络出版时间:2014-11-17

基金项目:国家“863”高技术发展计划项目(2012AA011804)

作者简介:廖荣华(1989-),男,硕士研究生,CCF会员,研究方向为智能交通系统、图形图像;兰时勇,助理研究员,博士,通信作者,研究方向为智能交通系统、图形图像;刘正熙,教授,研究方向为数字图像处理。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20141117.2209.042.html>

究中,基于混沌时间序列的预测方法直接从交通流的非线性特性入手进行预测,并且交通流混沌现象表面的随机性中蕴藏着系统内在秩序确定性,使得混沌系统在短期内的预测是可行的。

在文献[14]提出的混沌时间序列预测方法中,局域法比全域法效果好。文中在研究交通流混沌性的基础上,应用局域法中的加权零阶预测法和加权一阶预测法进行交通流的短时预测;同时针对算法流程中邻近相点的选取,采用欧氏距离和夹角向量两种选取方式。由于欧氏距离只能反映相点间的距离关系,于是文献[15]提出限制向量夹角的选取方式,通过相似演化来判断邻近相点。文中提出一种改进的邻近相点选取方法,利用各个相点与中心相的平均海明距离来反映距离关系;然后计算相点各个分量间的距离差和平均距离差,利用距离差之间的关系来表示相点的相似程度;最后将距离和相似程度综合起来判断邻近相点。将改进邻近点选取的预测方法和原方法进行对比,发现精确度更高。

1 交通流的混沌性

1.1 相空间重构

相空间重构可以将一维的时间序列构造出原系统的高维相空间结构。在高维相空间中,混沌动力系统的吸引子等许多重要的特性保存下来,方便研究动力学系统。1980年Packard等人提出导数重构法和坐标延迟重构法来重构相空间,现在一般采用坐标延迟重构法;1981年Takens证明嵌入定理,提出利用延迟时间和嵌入维数来重构相空间,重构的过程为:

设有单变量时间序列 $x(i), i=1, 2, \dots, N, N$ 是序列总长度。重构相空间为:

$$\begin{bmatrix} x(1) & x(2) & \dots & x(M) \\ x(1+\tau) & x(2+\tau) & \dots & x(M+\tau) \\ \vdots & \vdots & \vdots & \vdots \\ x[1+(m-1)\tau] & x[2+(m-1)\tau] & \dots & x[M+(m-1)\tau] \end{bmatrix}$$

其中, τ 为延迟时间; m 为嵌入维数。相空间中的相点表示为: $X_i = [x(i), x(i+\tau), \dots, x(i+(m-1)\tau)]$ 。其中, $i=1, 2, \dots, M, M=N-(m-1)\tau$ 。相空间重构关键在于选择合适的延迟时间 τ 和嵌入维数 m , 常用计算延迟时间 τ 的方法有自相关法、复相关法和互信息法等, 计算嵌入维数 m 的方法有关联维数法、虚假近邻法和 Cao 法等。

1.2 延迟时间和嵌入维数的选取

重构相空间延迟时间和嵌入维数的方法都认为两个参数各自独立, 分别求取。但另外一种观点认为两个参数存在某种相关性, 应该同时确定。1999年, Kim, Eykholt 和 Salas 提出了 C-C 算法, 利用关联积分

同时计算出最优延迟时间 τ 和嵌入窗宽 τ_w , 并利用公式 $\tau_w = (m-1)\tau$ 计算出嵌入维数 m 。

C-C 方法的具体步骤为:

步骤 1: 读入数据, 计算时间序列的标准差 σ , 根据统计结论, 取 $m=2, 3, 4, 5; r_j = i\sigma/2; i=1, 2, 3, 4; t=1, 2, \dots, 200$;

步骤 2: 将时间序列分拆成 t 个不相交的子序列, 长度 $l = [N/t]$:

$$\begin{aligned} &\{x(1), x(t+1), x(2t+1), \dots\} \\ &\{x(2), x(t+2), x(2t+2), \dots\} \\ &\vdots \\ &\{x(t), x(x+t), x(2t+t), \dots\} \end{aligned}$$

依据每个子序列, 分别计算公式(1)~公式(7): 关联积分:

$$C(m, N, r, t) = \frac{2}{M(M-1)} \sum_{1 \leq i < j \leq M} H(r - \|x_i - x_j\|) \quad (1)$$

统计量:

$$S(m, N, r, t) = C(m, N, r, t) - C^m(1, N, r, t) \quad (2)$$

其中, $H(r) = \begin{cases} 1, & r \geq 0 \\ 0, & r < 0 \end{cases}$ 为 Heaviside 阶跃函数, $M = N - (m-1)\tau$ 表示 m 维相空间的嵌入点数。

当 $N \rightarrow \infty$, 得到:

$$S(m, r, t) = \frac{1}{t} \sum_{s=1}^t [C_s(m, r, t) - C_s^m(1, r, t)] \quad (3)$$

$$\Delta S(m, t) = \max\{S(m, r_j, t)\} - \min\{S(m, r_j, t)\} \quad (4)$$

$$\bar{S}(t) = \frac{1}{16} \sum_{m=2}^5 \sum_{j=1}^4 S(m, r_j, t) \quad (5)$$

$$\Delta \bar{S}(t) = \frac{1}{4} \sum_{m=2}^5 \Delta S(m, t) \quad (6)$$

$$S_{cor}(t) = \Delta \bar{S}(t) + |\bar{S}(t)| \quad (7)$$

步骤 3: 根据求得的结果作出 $\bar{S}(t), \Delta \bar{S}(t), S_{cor}(t)$ 的图形, 最优延迟时间 τ 就选为 $\Delta \bar{S}(t)$ 的第一个极小值或 $\bar{S}(t)$ 的第一个零点; 嵌入窗宽 τ_w 选为 $S_{cor}(t)$ 的最小值, 利用公式 $\tau_w = (m-1)\tau$ 计算出嵌入维数 m 。

1.3 时间序列混沌性的判定

时间序列只有具有混沌特性时, 才能用混沌理论中的相关预测模型进行预测。最大 Lyapunov 指数是否大于零是判断时间序列是否满足混沌性的一种方法。计算最大 Lyapunov 指数有 wolf 方法、小数据量方法等。其中小数据量方法对小数据组可靠、计算量小、实时性好且容易操作。

首先对时间序列进行 FFT 变换, 计算时间序列平均周期 P ; 根据选取的延迟时间 τ 和嵌入维数 m , 重构

相空间;寻找混沌时间序列中每个相点 X_j 的最近邻点为 $X_{\eta(j)}$,并用时间序列平均周期 P 对邻近点的选取进行了短暂分离限制:

$$d_j(0) = \min_j \|X_j - X_{\eta(j)}\|, |j - \eta(j)| > P \quad (8)$$

然后对相空间中每个点 $X_{\eta(j)}$,计算该邻点对在 i 个离散时间步后的距离 $d_j(i)$:

$$d_j(i) = \|X_{j+i} - X_{\eta(j)+i}\| \quad (9)$$

其中, $i=1, 2, \dots, \min(M-j, M-\eta(j))$ 。

最后对每个 i ,求出所有 j 的 $\ln d_j(i)$ 平均 $y(i)$:

$$y(i) = \frac{1}{q\Delta t} \sum_1^q \ln d_j(i), d_j(i) \neq 0 \quad (10)$$

其中, Δt 是所用时间序列的样本周期; q 是非零 $d_j(i)$ 的数目。用最小二乘法做出回归线,该直线的斜率就是最大 Lyapunov 指数 λ 。

2 混沌预测模型

局域法中两种比较典型的预测算法是加权零阶预测法和加权一阶预测法。这两种局域预测法都是将相空间轨迹的最后一点作为中心点,把与中心点距离最近的若干轨迹点作为相关点,然后对这些相关点做出拟合,根据拟合结果再估计下一点的走向,最后在预测出的轨迹点的坐标中分离出所需要的预测值。

在加权零阶预测法和加权一阶预测法中,相空间中各点与中心点之间的空间距离是一个非常重要的参数,往往与中心点的空间距离最近的那几个点决定着预测的准确性。因此,为了在一定程度上提高预测的精度,将中心点的空间距离作为一个拟合参数引入预测过程,并且它还有一定的消噪能力。设中心点 X_M 的邻近点为 $X_{M_i}, i=1, 2, \dots, q$,且到 X_M 的距离分别为 d_i ; 设 d_m 是 d_i 中的最小值,定义各个邻近点 X_{M_i} 的权值为:

$$P_i = \frac{\exp(-k(d_i - d_m))}{\sum_{i=1}^q \exp(-k(d_i - d_m))} \quad (11)$$

其中, k 为参数,一般取 $k=1$; q 是进行拟合的邻近点的个数。

2.1 加权零阶局域预测法

加权零阶预测法是在零阶预测的方法基础上增加了权值,使得预测效果更好。

加权零阶局域法的具体算法简述为:

步骤 1: 数据处理与相空间重构;

步骤 2: 根据重构后的相空间,寻找中心点 X_M 的邻近点 $X_{M_i}, i=1, 2, \dots, q$;

步骤 3: 根据各个邻近点的权值,计算出 X_{M+1} :

$$X_{M+1} = \frac{\sum_{i=1}^q X_{M_i} e^{-k(d_i - d_m)}}{\sum_{i=1}^q e^{-k(d_i - d_m)}} \quad (12)$$

从 X_{M+1} 中得到第 m 个分量即为预测结果 $x(N+1)$ 。

2.2 加权一阶局域预测法

加权一阶局域预测法是用一阶线性拟合的方式来逼近相点的演化趋势,一阶局域线性拟合为:

$$X_{M_{i+1}} = a\mathbf{e} + bX_{M_i}, i=1, 2, \dots, q$$

其中, a, b 为拟合所需的实系数; \mathbf{e} 为一 q 维向量, $\mathbf{e} = (1, 1, \dots, 1)^T$; $X_{M_{i+1}}$ 是 X_{M_i} 演化一步后的相点。

当嵌入维数 $m=1$ 时,应用加权最小二乘法有

$$\sum_{i=1}^q P_i (x_{M_{i+1}} - a - bx_{M_i})^2 = \min \quad (13)$$

解上述方程组即可得到系数 a, b ,将 a, b 代入一步预测公式 $X_{M+1} = a\mathbf{e} + bX_M$, 即可得到演化一步后的相点预测值 X_{M+1} :

$$X_{M+1} = (x_{M+1}, x_{M+1+\tau}, \dots, x_{M+1+(m-1)\tau})$$

其中, X_{M+1} 中前 $(m-1)$ 个元素为原序列中已知值,其第 m 个元素 $x_{M+1+(m-1)\tau}$ 即为原序列的一步预测值 $x(N+1)$ 。

3 邻近点的选取

混沌预测模型中,邻近点的选取是个重要的工作。在重构的混沌相空间中,邻近的相点具有相似的演化行为,这是局域预测方法的基本前提和预测依据,目标预测点的运动趋势正是通过各邻近相点的运动趋势推断出来的,因此邻近相点的选取直接关系到预测的结果准确性。

3.1 欧氏距离

欧氏距离方法是利用相点间的距离关系:

$$d_i = \|X_i - X_M\| \quad (14)$$

直接选取 d_i 较小的 q 个值。

3.2 向量夹角

向量夹角方法通过限制相点演化轨迹间的夹角来选取邻近相点。向量夹角的具体步骤如下:

步骤 1: 根据欧氏距离在一定的阈值范围内,寻找离预测中心点 X_M 最近的 t 个初始邻近相点域 $X_{M_i}, i=1, 2, \dots, t$;

步骤 2: 采取往前回溯一步历史,即选取预测中心点 X_M 的上一个历史点 X_{M-1} ,以及 X_{M-1} 的最近邻点 $X_{\eta(M-1)}$,其中 $X_{\eta(M-1)}$ 的下一个演化点为 $X_{\eta(M-1)+1}$,用公式(15)来限制夹角,选取夹角小于 45 度的 q 个邻近点值,显然 $q \leq t$ 。

$$\cos\theta_i = \frac{(X_{M_i} - X_M) \cdot (X_{\eta(M-1)+1} - X_M)}{\|X_{M_i} - X_M\| \times \|X_{\eta(M-1)+1} - X_M\|} \quad (15)$$

3.3 改进方法

文中提出一种综合考虑相点距离和相似程度的方

法。首先计算各个相点与中心相点 X_M 的每个分量的差值： $x_{iMk} = x_{ik} - x_{Mk}$ ($k=1, 2, \dots, m; i=1, 2, \dots, M-1$)。

其中, m 是相空间的嵌入维数, M 是相点个数。并且得到每个分量的平均差值:

$$e_{iM} = \frac{1}{m} \sum_{k=1}^m x_{iMk} \quad (16)$$

然后计算两个相点间海明距离的平均值:

$$d_{iM} = \frac{A_{d_{iM}}}{m} = \frac{1}{m} \sum_{k=1}^m |x_{iMk}| \quad (17)$$

其中, d_{iM} 表示相点间的距离关系。

利用每个分量的差值与平均差值的关系来表示相点相似程度:

$$s_{iM} = \frac{1}{m} \sum_{k=1}^m |x_{iMk} - e_{iM}| \quad (18)$$

最后综合距离关系和相似程度来选取最近邻点:

$$C_{d_{iM}} = \frac{1}{2}(s_{iM} + d_{iM}) \quad (19)$$

$C_{d_{iM}}$ 越接近零, 表示选取的邻近相点效果越好。选取 $C_{d_{iM}}$ 最小的 q 个值表示 q 个最近邻点。

4 实验仿真

采用北京陶然亭公园路段 2008 年 6 月 17 日到 6 月 23 日七天每 5 min 的交通流量数据, 四个车道的数据。取这七天每天 6:30-18:30 的数据, 共 1 015 个数据作为样本, 预测第七天 18:35-21:00 这 30 个数据, 将得到的结果与实际测量值进行对比。

首先将初始样本值生成每 5 min 的交通流时间序列数据; 运用 C-C 算法计算时间序列得到最优延迟时间 $\tau = 27$, 嵌入窗宽为 $\tau_w = 30$, 则根据公式得到嵌入维数 $m = 3$; 然后根据选取的延迟时间和嵌入维数进行相空间重构, 恢复高维的相空间; 用小数据量法计算最大 Lyapunov 指数 $\lambda = 0.045\ 788 > 0$, 满足混沌特性, 可以运用预测模型进行预测; 最后利用加权零阶局域法和加权一阶局域法进行预测, 并与改进邻近点选取和权值计算后的两种方法的预测结果进行对比。

实验中选取的性能指标为:

平均绝对误差:

$$MAE = \frac{1}{N} \sum_{i=1}^N |f_i - y_i| \quad (20)$$

平均相对误差:

$$MRE = \frac{1}{N} \sum_{i=1}^N \frac{|f_i - y_i|}{y_i} \quad (21)$$

均方误差:

$$MSE = \frac{1}{N} \sqrt{\sum_{i=1}^N |f_i - y_i|^2} \quad (22)$$

图 1 和图 2 为三种选点方式预测值和实测值的对比图。

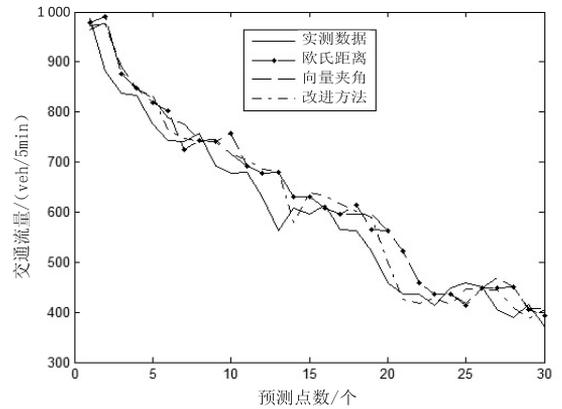


图 1 加权零阶预测法预测效果对比

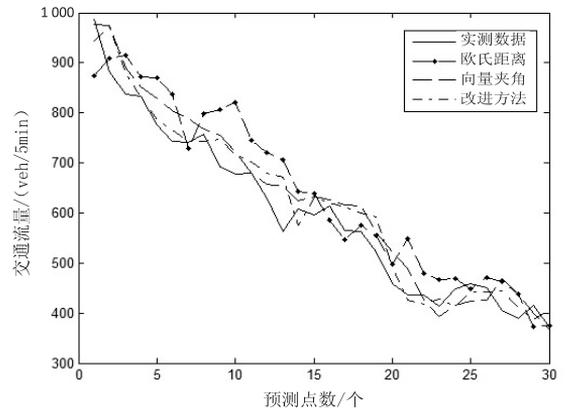


图 2 加权一阶预测法预测效果对比

从图中可以看出, 三种方法都能很好地预测, 其中改进后的方法比欧氏距离和向量夹角的预测值趋势更接近于实测值。

图 3 和图 4 为三种选点方式预测平均相对误差对比图。

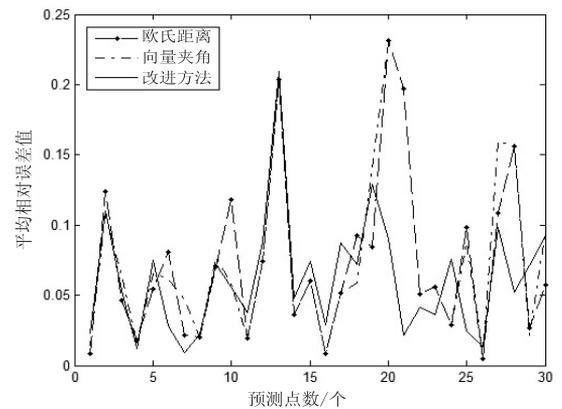


图 3 加权零阶预测法误差值对比

从图中可以看出, 改进后的方法预测误差较大的点比其他两种方法少, 并且预测的平均相对误差波动幅度比其他两种方法小。说明预测效果比欧氏距离和向量夹角方法好。

由公式(20)~公式(22)计算预测算法的预测误

差,见表 1。

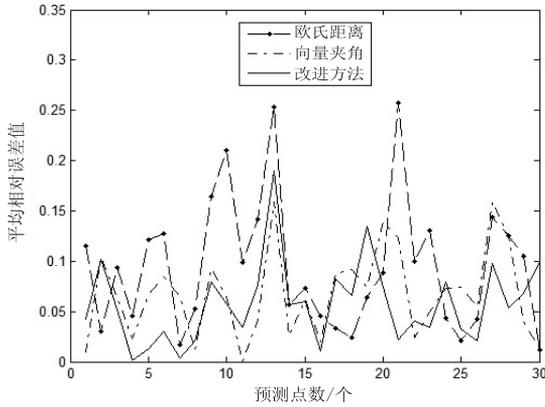


图 4 加权一阶预测法误差值对比

表 1 误差对比

预测方法	邻近点选取方式	MAE	MRE	MSE
加权零阶预测法	欧氏距离	41.021 4	0.073 6	9.391 3
	向量夹角	41.718 7	0.075 8	9.318 9
	改进方法	35.709 9	0.061 7	7.978 2
加权一阶预测法	欧氏距离	55.975 8	0.094 6	12.477 3
	向量夹角	38.366 0	0.067 5	8.242 5
	改进方法	33.047 3	0.057 8	7.452 3

从表 1 中可以看出,改进邻近点选取后,加权零阶局域法和加权一阶局域法的 MRE 都有明显的减少,同时 MAE 和 MSE 都有明显下降,说明改进后模型预测精度高于原模型。

5 结束语

交通系统复杂多变,交通流的预测方法多种多样。文中在混沌模型下研究短时交通流时间序列,采用常用的 C-C 方法得到重构参数,重构相空间并利用小数据量法得到最大 Lyapunov 指数判断交通流是否满足混沌性,然后采用混沌时间序列中的加权零阶局域预测法和加权一阶局域预测法来进行交通流预测。在局域法预测模型中,利用欧氏距离和夹角向量来选取邻近点,并且改进邻近点的选取,综合考虑相点间距离和相似程度。最后利用北京陶然亭公园路段 2008 年 6 月 17 日到 6 月 23 日七天的数据进行实验。实验结果表明,相比于初始预测算法,改进后算法的预测精度有了提高,说明改进后的预测算法能适用于短时交通流的预测,具有较高的精度。

参考文献:

- [1] 倪利华,陈笑蓉. ARIMA 模型结合小波去噪的贵阳市交通流预测[J]. 贵州大学学报:自然科学版,2011,28(5):87-91.
- [2] Ghosh B,Basu B,O'Mahony M. Multivariate short-term traffic flow forecasting using time-series analysis[J]. IEEE Transactions on Intelligent Transportation Systems,2009,10(2):246-254.
- [3] Han Chao,Song Su,Wang Chenghong. A real-time short-term traffic flow adaptive forecasting method based on ARIMA model[J]. Journal of System Simulation,2004,16(7):1530-1532.
- [4] 朱征宇,刘琳,崔明. 一种结合 SVM 与卡尔曼滤波的短时交通流预测模型[J]. 计算机科学,2013,40(10):248-251.
- [5] 张晓利,陆化普. 非参数回归方法在短时交通流预测中的应用[J]. 清华大学学报:自然科学版,2009,49(9):1471-1475.
- [6] 欧阳俊,陆锋,刘兴权,等. 基于多核混合支持向量机的城市短时交通预测[J]. 中国图象图形学报,2010,15(11):1688-1695.
- [7] Sapankevych N I,Sankar R. Time series prediction using support vector machines:a survey[J]. IEEE Computational Intelligence Magazine,2009,4(2):24-38.
- [8] 张秋余,朱学明. 基于 GA-Elman 神经网络的交通流短时预测方法[J]. 兰州理工大学学报,2013,39(3):94-98.
- [9] Xue Jieni, Shi Zhongke. Short-time traffic flow prediction based on chaos time series theory[J]. Journal of Transportation Systems Engineering & IT,2008,8(5):68-72.
- [10] 周金勇. 混沌时间序列预测模型研究[D]. 武汉:武汉理工大学,2009.
- [11] Iokibe T,Fujimoto Y,Kanke M. Short-term prediction of chaotic time series by local fuzzy reconstruction method[J]. Journal of Intelligent and Fuzzy Systems,1997,5(1):3-21.
- [12] 卓卉. 基于混沌理论的短时交通流预测方法研究[D]. 北京:北京交通大学,2008.
- [13] 董锐,贾元华,敖谷昌. 基于小波消噪和混沌时间序列的交通流预测[J]. 科学技术与工程,2010,10(31):7848-7851.
- [14] 张立,谢忠玉,陈凯. 基于混沌理论的短时交通流局域预测模型[J]. 黑龙江工程学院学报:自然科学版,2011,25(2):52-54.
- [15] 唐巍,谷子. 基于相关邻近点与峰谷荷修正的短期负荷时间序列预测[J]. 电力系统自动化,2006,30(14):25-29.

基于混沌时间序列局域法的短时交通流预测

作者: [廖荣华](#), [兰时勇](#), [刘正熙](#), [LIAO Rong-hua](#), [LAN Shi-yong](#), [LIU Zheng-xi](#)
作者单位: [四川大学 计算机学院, 四川 成都 610065](#); [四川大学 视觉合成重点学科实验室, 四川 成都 610045](#)
刊名: [计算机技术与发展](#) 
英文刊名: [Computer Technology and Development](#)
年, 卷(期): 2015(1)

引用本文格式: [廖荣华](#), [兰时勇](#), [刘正熙](#), [LIAO Rong-hua](#), [LAN Shi-yong](#), [LIU Zheng-xi](#) [基于混沌时间序列局域法的短时交通流预测](#)[期刊论文]-[计算机技术与发展](#) 2015(1)