

引用格式:殷礼胜,李胜,唐圣期,等. 基于改进的路网聚类结合 PSOWNN 的交通流预测[J]. 微电子学与计算机,2020,37(1): 20-26. [YIN L S, LI S, TANG S Q, et al. Traffic flow prediction based on improved road network clustering combined with PSOWNN[J]. Microelectronics & Computer, 2020, 37(1): 20-26.]

基于改进的路网聚类结合 PSOWNN 的交通流预测

殷礼胜, 李 胜, 唐圣期, 何怡刚

(合肥工业大学 电气与自动化工程学院, 安徽 合肥 230009)

摘 要: 针对城市路网交通流数据的空间相关性、非线性和平稳与非平稳的特点, 提出一种采用基于交通流量数据相关性分析改进的路网聚类算法与基于交通流量分段加权适应度函数的粒子群小波神经网络算法(MC-MP-SOWNN)相结合的预测方法, 来提高算法的预测精度. 首先, 利用基于交通流量数据相关性分析的路网聚类算法筛选出空间中与预测点交通流量数据相关系数高的其他观测点, 以此精简了样本输入数据, 减少冗余数据对预测精度的干扰, 提高整体预测精度; 其次, 再构建一种新型的粒子群算法的适应度函数, 给予整体预测样本中非平稳数据段更大的调节力度, 以此来进一步提高非平稳数据段的预测精度. 最后经实验结果分析, 提出的改进预测算法相比未进行改进前预测算法而言, 明显提高了整体及非平稳数据段预测精度, 达到较好的预测效果.

关键词: 交通流预测; 相关性路网聚类; 分段加权适应度函数; 粒子群算法; 小波神经网络

中图分类号: U491.1

文献标识码: A

文章编号: 1000-7180(2020)01-0020-07

Traffic flow prediction based on improved road network clustering combined with PSOWNN

YIN Li-sheng, LI Sheng, TANG Sheng-qi, HE Yi-gang

(School of Electrical Engineering and Automation, Hefei University of Technology, Hefei 230009, China)

Abstract: Aiming at the spatial correlation, nonlinearity and stationary and non-stationary characteristics of urban road network traffic flow data, this paper proposes a prediction method based on the improved traffic network data clustering algorithm based on traffic flow data correlation analysis and wavelet neural network algorithm based on traffic flow segmentation weighted fitness function particle swarm optimization algorithm (MC-MPSOWNN), to improve the prediction accuracy of the algorithm. Firstly, the road network clustering algorithm based on traffic flow data correlation analysis is used to screen out other observation points with high correlation coefficient between the traffic and the predicted traffic flow data, so as to simplify the sample input data and reduce the accuracy of the redundant data. Interference, improve overall prediction accuracy. Secondly, a new particle swarm algorithm fitness function is constructed to give greater adjustment of non-stationary data segments in the overall prediction sample, so as to further improve the prediction accuracy of non-stationary data segments. Finally, the experimental results show that the improved prediction algorithm proposed in this paper improves the convergence speed of the algorithm compared with the pre-predictive algorithm. It also improves the prediction accuracy of the whole and non-stationary data segments and achieves better prediction results.

Key words: traffic flow forecast; correlation road network clustering; improved particle swarm optimization; wavelet neural network

收稿日期: 2019-04-14; 修回日期: 2019-05-08

基金项目: 国家自然科学基金(51577046, 61673153); 国防科技计划项目(C1120110004, 9140A27020211DZ5102); 教育部科学技术研究重大项目(313018); 安徽省科技计划重点项目(1301022036)

1 引言

近年来,城市智能交通系统成为一个热点研究课题,而城市道路交通流预测对于城市交通系统能否提供准确高效的出行服务具有重要意义.各专家学者针对这一研究方向提出了众多预测模型.其中基于统计理论的模型有卡尔曼滤波、ARIMA 模型以及非参数回归模型等^[1-3],基于智能算法的预测模型有神经网络、支持向量机、模糊理论等^[4-6].而单一的预测模型已不能满足特定复杂交通环境中越来越高的精度要求,所以为了进一步考虑复杂交通环境中的交通流量影响因子,于是将交通网络中复杂周边交通环境及交通流量数据本身特性与智能算法相结合进行城市交通流预测成为一个重要研究领域.如芮兰兰等人^[7]通过对交通流时间序列数据的聚类 and 分割,再使用泛化能力强的极限学习机进行仿真,以保证数据之间更高的相似性,但数据来源单一其未考虑到交通流量预测中数据的空间特性.陆文琦等人^[8]将路网时空特性进行充分考虑后,再采用 ARIMA 模型和 PSO 优化的 BP 神经网络模型分别对交通流进行短时速度预测,得到最终的组合预测模型,但未考虑交通流数据间的相关性,忽视了数据冗余现象,进而影响模型整体的预测精度.邴其春等人^[9]提出通过灰色关联度分析确定交通流预测影响因子,然后采用粒子群优化算法构建非参数投影寻踪回归模型,但未进一步考虑到交通流数据非平稳特性,导致其在非平稳数据段未能达到更好的预测效果.

本文针对以上研究的不足在构建预测模型时,利用基于交通流量数据相关性分析的路网聚类算法搜集预测点周边符合一定交通空间特性且与预测点交通流量数据相关系数较高的观测点,剔除了相关系数低的冗余数据,精简样本数据的数量.再构建一种交通流量分段加权新型适应度函数给整体样本数据中非平稳数据段给予更大的调节力度,以此进一步来提高非平稳数据段的预测精度.

2 基于交通流量数据相关性分析改进的路网聚类算法

在路网主要交通道路中随机设置交通流量数据观测点,本文聚类算法的改进在于将各观测点的交通流数据与预测点的交通流数据间的相关系数作为路网聚类依据之一.首先通过设置凝聚度 α 、扩展度 β 作为聚类条件,使路网的研究限制在预设路网实

际距离范围内,舍去范围之外的点,从而保证了路网空间实际距离的关联性,同时再在此基础上,考虑路网聚类中的交通流量数据间的相关性,精简了样本输入数据.

本文取 $[\alpha, \beta]$ 范围内与预测点交通流量数据相关系数大于 0.7 的所有观测点,完成聚类 C (见 4.2 节).为简化交通网络复杂性,将路网中的路口看作一个节点,而节点与节点间的路段作为一段弧.因此一个路网可以由一个无向带权图 $G = (V, E, W)$ 来表示,其中 V 为各路口节点集合, E 表示各路段弧的集合, W 为各段弧权值的集合.

路网中观测点分布在各条路段上,例如 $p_i = (u_i, u_j, \text{dis})$ 表示观测点 p_i , 其中 u_i 与 u_j 表示为观测点所在路段最近的两个路口节点, dis 为 p_i 与 u_i 的距离(单位为米).此处 i, j 用以表示区分 G 中不同点的下标.

2.1 路网聚类中各点实际路网距离及参数的定义与界定

2.1.1 路网聚类中各点实际路网距离的定义

(1) 同一路段弧上两个观测点间的实际路网距离 d_D ;

$$d_D(p, q) = |\text{dis}_p - \text{dis}_q| \quad (1)$$

例如点 $T_1 = (A, B, 56)$ 见图 1 所示.

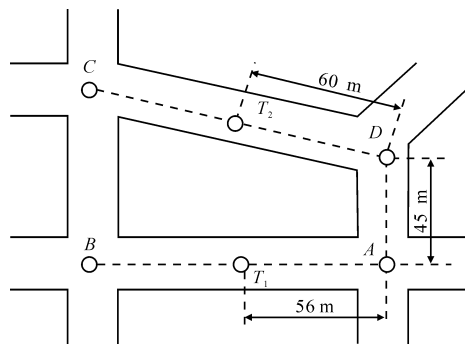


图 1 路网中点与路网的关系示意图

(2) 路口节点之间的实际路网距离. $d_D(u_i, u_j)$ 为路口节点 u_i 与 u_j 间的最短实际路网距离, 由 Dijkstra 算法^[10] 算得. 如图 1 所示节点 D 与节点 A ($d_D(D, A) = 45$).

(3) 不同路段弧上观测点间实际路网距离.

$$d_D(p, q) = d_D(p, u_i) + d_D(u_i, u_j) + d_D(u_j, q) \quad (2)$$

如 $T_1 = (A, B, 56)$ 与 $T_2 = (D, C, 60)$, 如图 1 所示, 其距离 $d_D(T_1, T_2)$ 为 T_2D 段、 DA 段和 AT_1 段距离之和 ($d_D(T_1, T_2) = d_D(T_2, D) + d_D(D, A) + d_D(A, T_1)$).

2.1.2 基于交通流数据相关性路网聚类中参数取值的界定

设 p_i 为预测点,则凝聚度与扩展度取值的界定如下:

凝聚度 α 值为与预测点实际路网距离最近的路口节点间的实际路网距离,见式(3):

$$\alpha = \min\{d_D(p_i, u_j)\}, u_j \in \mathbf{G} \quad (3)$$

扩展度 β 值的选取规则为与预测点最相邻的连续两个路口的实际路网距离之和,如式(4):

$$\beta = \min\{d_D(p_i, u_j)\} + \min\{d_D(u_j, u_k)\}, \\ u_j, u_k \in \mathbf{G} \quad (4)$$

式中, $\min\{d_D(u_j, u_k)\}$ 表示 $\mathbf{G}$ 中与 u_j 最近的路口节点的实际路网距离,本文研究路网中具体取值见 4.1 节. 此处 i, j, k 表示 $\mathbf{G}$ 中不同节点下标.

2.2 基于交通流量数据相关性改进的路网聚类算法步骤

通过以上对无向带权图 $\mathbf{G}$ 的构建及交通路网中各点与参数的定义与界定,以下为改进的路网聚类算法的具体步骤.

步骤 1:选取路网中待研究的预测点,设预测点 $p_i = (u_a, u_b, \text{dis}_i)$,创建集合 $\mathbf{Q}$,将预测点加入 $\mathbf{Q}$,由公式(3)和(4)根据 $\mathbf{G}$ 中路网各路口实际路网距离设定凝聚度 α 、扩展度 β ;

步骤 2:由预测点 p_i 出发,分别向 u_a, u_b 方向各自进行搜索至节点 u_a, u_b ;设 p_k 为预测点 p_i 在节点 u_b 方向上的邻接观测点;

(1) 若 p_k 存在,判断是否有 $\alpha \leq d_D(p_k, p_i) \leq \beta$,若符合,则将点 p_k 加入 $\mathbf{Q}$,继续向 u_b 方向扩展遍历至节点 u_b ;

(2) 若 p_k 不存在,则进行步骤 3;

步骤 3:遍历节点 u_b 的所有邻接节点,设 u_b 有邻接节点 j 个 $(u_c^1, u_c^2, \dots, u_c^j)$,其中 j 为大于或等于 1 的正整数. 设 p_k 为路段 (u_b, u_c^j) 上邻接 u_b 的观测点,设当前 $j = 1$;

(1) 若存在 p_k ,判断条件 $\alpha \leq d_D(p_k, p_i) \leq \beta$,若满足则加入 $\mathbf{Q}$ 中,继续向 u_c^1 方向扩展至点 u_c^1 ,直到遍历所有观测点;

(2) 若不存在 p_k ,则直接进行步骤 4;

步骤 4:进行 $j = j + 1$ 操作,同理继续判断路段 (u_b, u_c^j) 上邻接 u_b 的观测点,直到节点 u_c^j ;判断 j 是否达到最大值,若是,则执行步骤 5,否则继续执行步骤 4;

步骤 5:分别发散所有 $(u_c^1, u_c^2, \dots, u_c^j)$ 节点的邻接节点,以此类推,直到聚集所有满足条件 $\alpha \leq$

$d_D(p_k, p_i) \leq \beta$ 的点;预测点 p_i 在 u_a 方向上的发散同理;

步骤 6:将以上步骤中挑选出来的 $\mathbf{Q}$ 中所有观测点的历史交通流量数据与预测点的历史交通流量数据作相关性分析;选择与预测点交通流量数据相关系数大于 0.7 的所有观测点,将其划分为一类路网聚类空间中,完成聚类 $\mathbf{C}$,具体聚类结果见 4.2 节.

3 基于分段加权适应度函数改进的 PSOWNN 预测算法

3.1 小波神经网络

小波神经网络是一种以 BP 神经网络拓扑结构为基础,将小波基函数作为隐含层节点的激活函数,信号前向传播同时误差反向传播的神经网络,具有结构简单和更强的学习能力的优点. 本文采用小波神经网络为研究工具.

3.2 粒子群算法

粒子群算法中每个粒子都通过跟踪个体极值 pbest 和群体极值 Gbest 来更新自身位置,每更新一次位置时计算各粒子的适应度,再根据适应度来更新 pbest 和 Gbest 的值,不停迭代直到满足条件.

例如,在每次迭代中,第 i 个粒子通过个体极值和全局极值来更新自身的速度和位置:

$$v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (p_{id}^k - z_{id}^k) + c_2 r_2 (p_{gd}^k - z_{id}^k) \quad (5)$$

$$z_{id}^{k+1} = z_{id}^k + v_{id}^{k+1} \quad (6)$$

式中, $i = 1, 2, \dots, m; d = 1, 2, \dots, D; k$ 是迭代次数; v_{id} 为粒子的速度矢量元素; z_{id} 为粒子的位置矢量中的元素; p_{id} 为粒子个体极值矢量元素, p_{gd} 为种群极值矢量元素. r_1 和 r_2 是 $[0, 1]$ 之间的随机数; c_1 和 c_2 为学习因子,通常取 $c_1 = c_2 = 1.5$,取非负常数;为防止粒子的盲目搜索,粒子的位置和速度通常分别限制在区间 $[-5, 5]$ 和 $[-1, 1]$ 内. ω 为惯性权重,通常的取法为 $\omega_{\max} = 0.9, \omega_{\min} = 0.4$,起着权衡局部最优能力和全局最优能力的作用.

3.3 基于交通流量分段加权的新型适应度函数的构建

将交通流量训练样本输出数据与交通流量实际值按时间依次统一均等地划分为 N 个数据段;其次,分别求解各自每一个数据段的交通流量实际值的方差值,以此来表征不同段的非平稳情况. 其样本数据分段示意如图 2.

基于交通流量各数据段加权适应度函数 F 的构建方法见式(7) ~ (10):

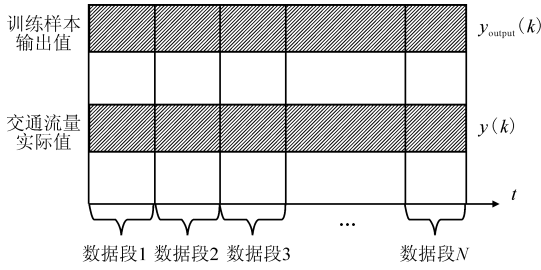


图 2 交通流量数据分段示意图

$$s^2 = s_1^2 + s_2^2 + \dots + s_N^2 \quad (7)$$

$$E_i = \sum_{k=(i-1) \times l+1}^{i \times l} |y_{\text{output}}(k) - y(k)|, \quad i = 1, 2, \dots, N; \quad (8)$$

$$f(s^2, s_1^2, \dots, s_N^2, E_1, \dots, E_N) = \frac{s_1^2}{s^2} E_1 + \frac{s_2^2}{s^2} E_2 + \dots + \frac{s_N^2}{s^2} E_N \quad (9)$$

$$F = 1/f(s^2, s_1^2, \dots, s_N^2, E_1, \dots, E_N) \quad (10)$$

式中, $s_1^2, s_2^2, \dots, s_N^2$ 分别是交通流量实际值各段的方差值; s^2 是各段交通流量方差值之和; n 为训练样本数(本文实验中 $n = 288$, 见 4.1 节). 为了使每段能够充分反映整个时间序列的平稳与非平稳情况, 取 $N \approx n/l$, 其中 l 为交通流量样本数据中最短的非平稳时间序列的长度. 其中最短的非平稳时间序列界定是由非平稳时序数据的分段辨识及其递推算法^[11]得出, 该算法通过分析时间序列中的各子时序的均值、方差和自相关函数的数字特征来析出平稳与非平稳序列的分界点, 以此得到各非平稳序列. 经界定本文交通流时间序列 $y(k), k = 1, 2, \dots, n$ 中的最短非平稳时间序列为 $y(8) \sim y(20)$ 数据段, 故取 $l = 20 - 8 = 12, N = n/l = 288/12 = 24; E_1, E_2, \dots, E_N$ 分别是各数据段训练样本输出数据与实际值之间的误差和, $y_{\text{output}}(k)$ 为训练样本输出值, $y(k)$ 为交通流量实际值; $f(s^2, s_1^2, \dots, s_N^2, E_1, \dots, E_N)$ 为经过不同段方差值与各段方差值总和之比加权后的误差和函数, F 为 $f(s^2, s_1^2, \dots, s_N^2, E_1, \dots, E_N)$ 的倒数, 以此作为粒子群算法中的适应度函数.

通过不同数据段内交通流量的方差来表征不同段间的各自非平稳状况, 方差值 $s_i^2 (i = 1, 2, \dots, N)$ 越大, 则说明第 i 数据段中交通流数据表现出明显的波动趋势, 表现出非平稳特性. 而这些数据段往往表现出更强随机性, 预测难度大, 预测的精度好坏更加关系到整个预测模型整体预测效果.

基于传统的整体训练误差和的适应度函数, 无法针对非平稳数据段给予更强的调节力度, 针对性不强, 导致整体以及非平稳数据段预测精度低的现

象, 而本文构建的交通流量分段方差加权适应度函数, 根据各段非平稳情况, 分配不同权重, 提高对非平稳数据段的调节力度, 改善非平稳数据段预测难度大的缺陷.

将小波神经网络中输入层到隐含层连接权值 w_{ij} 、隐含层到输出层连接权值 w_{jk} 、平移因子 b_j 以及伸缩因子 a_j 作为粒子位置向量中的各个元素(此处的 i, j, k 分别表示小波神经网络的输入层节点、隐含层节点、输出层节点), 让粒子群算法在迭代寻优过程中, 找到使其适应度最大的粒子位置向量, 完成寻优过程^[12-13].

3.4 基于交通流量分段加权适应度函数改进的 PSOWNN 预测算法步骤

基于以上对交通流数据进行时间序列均等分割, 再由各数据段方差值进行权重分配的新型适应度函数构建的粒子群小波神经网络, 其具体算法步骤如下:

步骤 1: 将路网聚类空间 C 中的所有观测点在 $t-1$ 时刻的交通流量数据序列 $\{Q_1(t-1), \dots, Q_M(t-1)\}$ 依次作为系统的前 M 个空间序列输入, 将预测点自身的 $t-1$ 时刻交通流量数据 $Q_{M+1}(t-1)$ 作为系统中的第 $M+1$ 个输入, 以此来作为预测 t 时刻交通流量的系统输入序列.

步骤 2: 根据前面路网聚类结果, 聚类 C 中有三个观测点(见 4.2 节), 根据得到的观测点数量 $M = 3$, 来设定粒子群小波神经网络输入层节点个数 $M+1 = 4$ 、隐含层节点依经验取 6、输出层节点为 1, 构建小波神经网络拓扑结构为 4-6-1;

步骤 3: 一般取粒子种群规模为 30, 每个粒子的维度为神经网络中需要迭代寻优参数的个数. 将神经网络中各权值 w_{ij} (输入层到隐含层)、 w_{jk} (隐含层到输出层)、平移因子 b_j 以及伸缩因子 a_j 作为每个粒子的位置向量中的元素, 由小波神经网络拓扑结构 4-6-1 可知 w_{ij}, w_{jk}, b_j, a_j 数量为 $4 \times 6 + 6 + 6 + 6 = 42$, 所以设置每个粒子维度为 42; 对粒子的速度与位置分别进行初始化, 其范围一般限定为 $[-1, 1], [-5, 5]$ 之间的随机数, 设置最大迭代次数 100;

步骤 4: $N = 24$ (由 3.3 节可知), 定义适应度函数 $F = 1/f(s^2, s_1^2, \dots, s_{24}^2, E_1, \dots, E_{24})$, 其中 $f(s^2, s_1^2, \dots, s_{24}^2, E_1, \dots, E_{24}) = \frac{s_1^2}{s^2} E_1 + \frac{s_2^2}{s^2} E_2 + \dots + \frac{s_{24}^2}{s^2} E_{24}$;

步骤 5: 输入数据经过归一化后作为预测系统的输入样本, 经过小波神经网络训练, 得出训练样本输出 $y_{\text{output}}(k)$, 再经过数据反归一化处理后, 计算

$E_1, E_2, \dots, E_{24}$.

步骤 6: 计算各粒子适应度值, 更新个体极值与群体极值, 更新粒子位置 (w_{ij}, w_{jk}, b_j, a_j) 和速度, 再判断是否达到最大迭代次数 100? 若是, 执行步骤 7; 若没有则重复本步骤 6, 直到满足条件为止;

步骤 7: 由以上步骤得到最优初始权值 w_{ij} (输入层到隐含层)、 w_{jk} (隐含层到输出层)、平移因子 b_j 及伸缩因子 a_j ; 再进行神经网络训练并计算训练误差 $e = y_{\text{output}}(k) - y(k)$, 其中 $k = 1, 2, \dots, n, y(k)$ 为交通流量实际值, 再由误差反馈调整 w_{ij}, w_{jk}, b_j, a_j ;

步骤 8: 设置训练最大次数 100, 判断迭代次数是否达到最大值 100 或平均绝对误差小于 3? 若没有, 则返回步骤 7, 否则停止训练. 再利用训练好的网络进行仿真预测, 得到网络预测结果.

其算法流程如图 3.

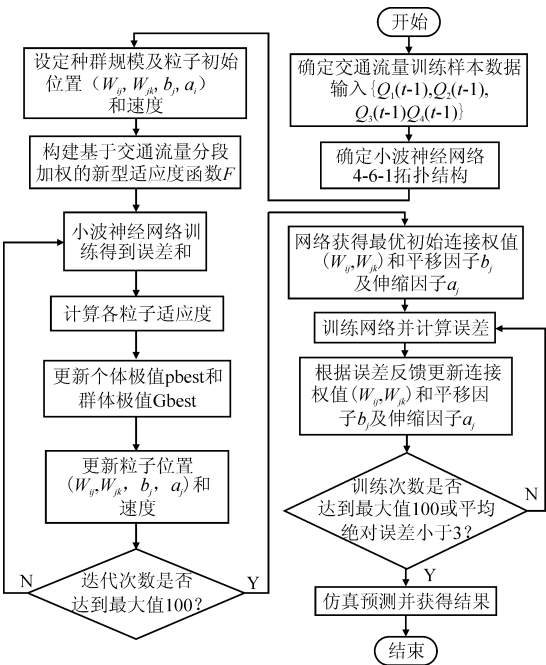


图 3 基于交通流量分段加权适应度函数改进的 PSOWNN 预测算法流程图

4 实例分析

4.1 交通流量数据来源

本文选定合肥市徽州大道与长江中路南交叉进口处为预测点, 取样时间为 2018 年 8 月 20 日至 24 日的工作日, 对其研究选择 6:00 到 12:00 时段 (此时段已包含交通流量数据的空间相关性、非线性和平稳与非平稳的特点), 对其进行实地观测并记录数据. 经该路口各观测点与预测点间道路交通实际距离的测算, 设置此路段凝聚度 $\alpha = 55$ 、扩展度 $\beta = 108$, 具体测算示意图见图 5. 通过路网聚类算法, 以

预测点为聚类中心, 沿路网方向四周搜寻符合路网距离 [55 m, 108 m] 内的所有观测点, 得到路网聚类中除预测点外其他的 5 个观测点. 对每个观测点同样进行上述时间交通流量数据的采集, 每 5 min 为一个统计间隔, 共 360 组交通流量数据, 前四天 288 组数据作为网络训练样本, 第五天 72 组数据作为预测样本. 路网卫星图见图 4.



图 4 徽州大道与长江中路交叉口卫星图

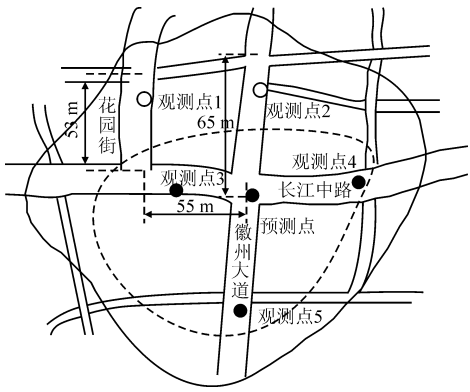


图 5 徽州大道与长江中路交叉口处各观测点位置平面图

4.2 基于交通流量观测点数据相关性分析

定义相关系数分析表达式:

$$\rho_{XY} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)} \sqrt{E(Y^2) - E^2(Y)}} \quad (11)$$

式中, X 与 Y 分别是观测点和预测点的交通流量序列, ρ_{XY} 为观测点与预测点间交通流量时间序列的相关系数, $E(X)$ 表示序列 X 的数学期望^[14].

取观测点 1、观测点 2、观测点 3、观测点 4 以及观测点 5 与预测点按式 (11) 进行数据相关性分析, 计算得到的各点相关系数见表 1.

表 1 各观测点与预测点交通流量数据相关系数

观测点序号	1	2	3	4	5
ρ_{XY}	0.442	0.677	0.831	0.848	0.900

一般来说, 相关系数取绝对值后, 0~0.09 为没

有相关性,0.1~0.3 为弱相关,0.3~0.5 为中等相关,0.5~1.0 为强相关.本文考虑到精简训练样本数据,此处选择大于 0.7 各观测点作为训练样本.由此可见,观测点 1 和观测点 2 与预测点数据相关系数均小于 0.7,故舍去;观测点 3、观测点 4、观测点 5 与预测点交通流量数据相关系数大于 0.7,使其聚为一类 **C** 中.作为聚类结果,即作为数据训练样本,见图 5 中虚线圈内所示.

4.3 交通流量曲线预测精度评价指标

对于各模型预测精度评价指标见如下方法:

(1)平均绝对误差

$$MAE = \frac{1}{n} \sum_{t=1}^n |P - P_r|$$
 (12)

(2)均方根误差

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (P - P_r)^2}$$
 (13)

式中, P 为实验所得预测结果; P_r 为真实交通流量数据.

4.4 实验预测结果及预测精度误差分析

本文分别使用经过未改进路网聚类结合未改进粒子群小波神经网络(C-PSOWNN)、未改进路网聚类结合改进粒子群小波神经网络(C-MPSOWNN)、改进路网聚类结合未改进粒子群小波神经网络(MC-PSOWNN)以及改进的路网聚类结合改进的粒子群小波神经网络(MC-MPSOWNN)的四种预测模型进行仿真相比较,各模型其预测情况如图 6~图 9 所示.

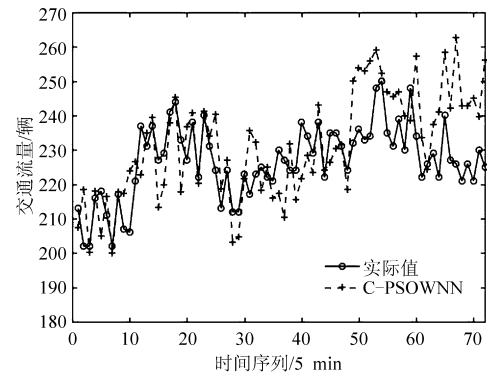


图 6 C-PSOWNN 模型预测示

由各模型预测示意图可知,图 6 未做改进的模型(C-PSOWNN)在时间序列非平稳区间[8,20]与区间[55,72]的预测效果差,很难准确反映真实交通情况,原因是没有进行交通流数据的有效筛取,导致数据间的相关性不高,无法排除冗余数据对预测精度的干扰.也没有对交通流数据进行分段加权分配

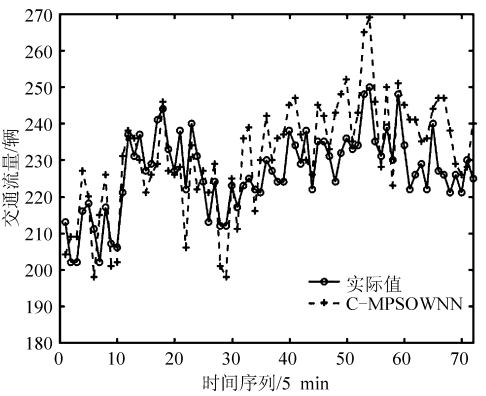


图 7 C-MPSOWNN 模型预测示意图

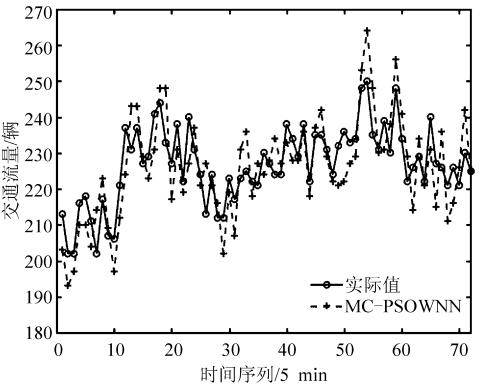


图 8 MC-PSOWNN 模型预测示意图

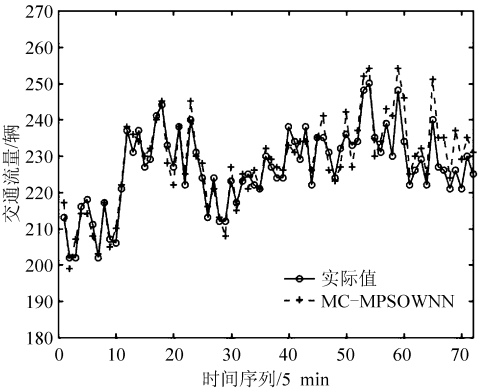


图 9 MC-MPSOWNN 模型预测示意图

各段训练调节力度,进而没能提高对非平稳期预测的精度.图 7 进行交通流量分段加权适应度函数的改进后与图 6 相比较而言,在非平稳期的预测精度明显改善,图 8 中模型经过改进的路网聚类算法后,提高了数据间的交通相关性,其预测效果也明显较图 6 有所较大改善,在非平稳期明显提高了预测精度达到较好的预测效果,但在非平稳期[8,20]和[55,72]区间内都仍有明显跳跃现象.而图 9 中经过基于交通流量数据相关性分析的路网聚类算法和基于交通流量分段加权适应度函数改进后粒子群小波神经网络(MC-MPSOWNN)两种算法的结合后,其

模型在非平稳期的预测达到最好的效果,整体上的预测精度也明显得到改善.

对于各预测模型在非平稳期数据段[8,20]与[55,72]区间的曲线预测误差指标如表 2.

表 2 各预测模型在非平稳期误差指标分析				
误差指标	MC-MPSO WNN	MC-PSO WNN	C-MPSO WNN	C-PSO WNN
MAE	3.214 1	7.457 1	9.342 2	10.057 4
RMSE	3.828 6	8.297 8	10.486 1	13.274 3

各预测模型整体预测误差评价指标情况如表 3.

表 3 各预测模型整体误差指标分析				
误差指标	MC-MPSO WNN	MC-PSO WNN	C-MPSO WNN	C-PSO WNN
MAE	2.986 1	7.097 2	9.027 8	9.527 8
RMSE	3.450 1	8.070 6	10.432 7	11.537 4

由误差指标可见,经过改进的路网聚类算法结合基于交通流量分段加权适应度函数改进后的粒子群小波神经网络预测方法,无论在非平稳数据段还是在整体预测效果上其平均绝对误差 MAE 和均方根误差 RMSE 均优于其他未改进模型的预测效果.

5 结束语

本文采用基于交通流量数据相关性改进的路网聚类算法结合基于交通流量分段加权适应度函数改进的粒子群小波神经网络的方法来进行交通流量预测,精简了样本数据,排除了相关系数低的冗余数据对预测精度的干扰,提高了算法整体预测精度.通过构建新型适应度函数,加强了对非平稳数据段的调节力度,提高了非平稳数据段的预测效果.最后通过实例分析验证了本文的预期,改进的模型在整体及非平稳段都具有更高的预测精度,更好满足了交通流预测的精确性的要求.该文在后续的工作中将进一步探讨和分析不同节假日与天气情况下的不同模型间预测的精度差异,以便进一步提炼交通预测的特殊影响因子,找到更加优化的解决方案.

参考文献:

[1] 罗赞文,吴志坚,韩曾晋. 扩展 Kalman 滤波法在高速公路交通流模型上的应用[J]. 自动化学报,2002(1): 90-96.

[2] 成云,成孝刚,谈苗苗,等. 基于 ARIMA 和小波神经网络

组合模型的交通流预测[J]. 计算机技术与发展, 2017,27(1):169-172.

[3] 冯青平,李星毅. 基于大数据的非参数回归短时交通流预测方法[J]. 无线通信技术, 2015, 24(3):38-43.

[4] 赵忠杰,田梅,王英伟. 基于神经网络的隧道小时交通量预测方法[J]. 微电子学与计算机, 2007, 24(11): 134-137.

[5] 梅朵,郑黎黎,刘春晓,等. 基于混合算法优化 SVM 的短时交通流预测[J]. 计算机技术与发展, 2017, 27(11):92-95.

[6] 程山英. 基于模糊神经网络的短时交通流预测方法研究[J]. 计算机测量与控制, 2017,25(8):155-158.

[7] 芮兰兰,李钦铭. 基于组合模型的短时交通流量预测算法[J]. 电子与信息学报, 2016, 38(5):1227-1233.

[8] 陆文琦,谷远利,陈伦. 基于时空融合的城市快速路短时交通流预测[J]. 计算机仿真, 2018,35(9):136-140.

[9] 邴其春,龚勃文,林赐云,等. 基于粒子群优化投影寻踪回归模型的短时交通流预测[J]. 中南大学学报(自然科学版), 2016,47(12):4277-4282.

[10] 王华. 改进 Dijkstra 算法的城市道路最短路径仿真研究[J]. 测绘科学, 2013,38(4):149-151.

[11] 黄雄波. 非平稳时序数据的分段辨识及其递推算法[J]. 计算机系统应用, 2017,26(5):180-185.

[12] DENG C, WANG F, SHI H M, et al. Real-Time free-way traffic state estimation based on cluster analysis and Multiclass Support Vector Machine[C]//2009 International Workshop on Intelligent Systems and Applications. Wuhan, China, 2009:1-4.

[13] ZHU G Y, CHEN J J, ZHANG P. Fuzzy C-means clustering identification method of urban road traffic state[C]//12th International Conference on Fuzzy Systems and Knowledge Discovery. Zhangjiajie, China, 2015:302-307.

[14] 肖尚华,胡灿林. 基于加权 K-means 聚类与路网无向图的地图分割算法[J]. 现代计算机(专业版), 2018,34(8):78-81.

作者简介:

殷礼胜 男,(1974-),硕士生导师,副教授.研究方向为复杂系统建模、交通流预测等.

李 胜(通讯作者) 男,(1993-),硕士研究生.研究方向为交通流预测、智能控制技术. E-mail:lishengzx@163.com.

唐圣期 男,(1995-),硕士研究生.研究方向为交通流预测、智能控制系统.

何怡刚 男,(1966-),博士生导师,教授.研究方向为复杂电磁分析与建模等.