

改进的 LS - SVM 数学模型的交通流量预测分析

吴一凡

(江苏食品药品职业技术学院基础教学部, 江苏 淮安 223003)

摘 要:针对交通流量数据具有非线性和非平稳性的特点,运用 EMD 和 FOA 算法实现 LS - SVM 核参数和惩罚系数的自适应优化选择,提出了一种基于 EFLS - SVM 算法的交通流量预测模型。通过 EMD 提取交通流量数据的细节特征和趋势特征,构建出基于 EFLS - SVM 的交通流量预测模型,分别进行单步、3 步、5 步和 7 步预测。通过不同交通流量预测模型的实验对比发现,EFLS - SVM 算法的预测精度和预测效率均优于其他模型,从而为交通网络资源的合理配置提供科学决策的依据。

关键词:交通流量;果蝇优化算法;数学模型;最小二乘法支持向量机

中图分类号:TP391.1

文献标志码:A

引 言

随着城市化规模和交通网络的日益增大,导致交通网络管理工作的繁重程度急剧上升,交通事故频发。为了保证交通资源的合理分配,高质量的交通流量预测对交通网络的规划、管理和设计具有重要的理论意义和实际价值。

雷霆^[1]等人将小波变换理论和神经网络技术相结合提出一种基于小波神经网络的交通流量预测模型,通过小波变换提取交通流量的细节特征和整体特征,将提取出来的交通流量特征作为神经网络的输入,并将实际交通流量作为输出,仿真结果表明预测精度较高,但其稳定性有待于进一步提高。

Guo Wen^[2]等人针对 BP 神经网络局部最优的问题,运用 PSO 算法的全局寻优的优点对 BP 网络进行改进,并将其应用于交通流量预测,改进后模型的预测精度和收敛速度优于传统的 BP 神经网络,缺点是 PSO 存在早熟问题,容易陷入局部最优。

刘渊^[3]等人将混沌理论引入小波神经网络进行网

络改进,仿真结果说明混沌小波神经网络的预测误差远远小于 RBF 神经网络的交通流量预测,但对小波基函数的选择和确定难度较大。

杨光^[4]等人结合小波核函数的多分辨率特性,提出一种基于小波核 LS - SVM 的交通流量预测,通过实际数据的预测验证,说明该方法具有一定的优越性,但参数的选择需要靠经验确定。

针对交通流量数据的非平稳和非线性的特点,本文结合 EMD 和 FOA 算法对 LS - SVM 核参数和惩罚系数进行自适应优化,提出一种基于 EFLS - SVM 算法的交通流量预测模型。通过 EMD 提取交通流量的细节特征和趋势特征,构建出基于 EFLS - SVM 的交通流量预测模型的输入和输出,实现交通流量的预测,为交通网络资源的优化配置提供科学合理决策的依据。

1 果蝇优化算法

果蝇优化算法(Fruit Fly Optimization Algorithm, FOA)由潘文超^[5]于 2011 年提出的一种全新的演化式计算方法。该算法具有控制参数少、收敛速度快和收敛

收稿日期:2015-10-19

作者简介:吴一凡(1980-),男,江苏淮安人,讲师,硕士,主要从事数学模型算法等方面的研究,(E-mail) wuyifan1124@sina.com

精度高的优点,目前被广泛地应用于函数最优化、SVM参数优化、背包问题以及神经网络权值和阈值的优化等。

果蝇优化算法具体流程如下^[5]:

(1) 设置果蝇优化算法的果蝇群体大小 $popsiz$ 和最大迭代次数 $Iteration$, 随机初始化果蝇群体位置, 初始化结果分别用 X_begin 和 Y_begin 表示。

(2) 根据公式(1)和公式(2)计算果蝇个体进行寻优的随机方向和距离;

$$x_i = X_begin + Value \times rand() \quad (1)$$

$$y_i = Y_begin + Value \times rand() \quad (2)$$

其中, $Value$ 表示果蝇的搜索距离; x_i 和 y_i 分别表示果蝇个体的下一时刻的位置。

(3) 根据式(3)估计果蝇个体和原点之间的距离 d_i , 之后运用式(4)计算果蝇个体的味道浓度 s_i :

$$d_i = \sqrt{x_i^2 + y_i^2} \quad (3)$$

$$s_i = \frac{1}{d_i} \quad (4)$$

(4) 味道浓度 s_i 代入式(5)味道浓度判定函数, 计算出该果蝇个体当前位置的味道浓:

$$Smell_i = Function(s_i) \quad (5)$$

(5) 找到果蝇群体中最佳味道浓度值和最佳位置, 最佳味道浓度由 $Smell_b$ 表示, 最佳位置由 x_b 和 y_b 表示。

(6) 保留并记录果蝇最佳位置和最佳味道浓度, 最佳味道浓度 $Smell_{best} = Smell_b$, 果蝇初始位置 $X_begin = x_b$, $Y_begin = y_b$, 同时果蝇群体朝着该最佳位置搜寻过去。

(7) 进入迭代寻优, 重复迭代步骤(2)~(5), 同时判断味道浓度是否好于前一迭代味道浓度; 若成立, 则执行步骤(6)。

2 经验模态分解

经验模态分解吸取了小波变换多分辨的优点, 同时克服了小波变换中小波基选择和分解尺度很难确定的缺点, 因此 EMD 非常适合分析交通流量序列, 因为交通流量序列是非线性、非平稳序列。

交通流量时间序列的 EMD 分解过程可分为^[6]:

(1) 识别交通流量时间序列数据中的所有极大值点, 并计算拟合出其上包络线 $e_{up}(t)$ 。

(2) 提取交通流量时间序列数据中的极小值点, 并计算拟合出其下包络线 $e_{low}(t)$, 在上下包络线的基础上, 计算上下包络线的平均值 $m_1(t)$

$$m_1(t) = \frac{e_{up}(t) + e_{low}(t)}{2} \quad (6)$$

(3) 将交通流量数据序列 $x(t)$ 减去 $m_1(t)$, 得到 $h_1(t)$, 将 $h_1(t)$ 看成新的交通流量数据 $x(t)$, 重复步骤(1), 经过 k 次筛选, 直至 $h_1(t) = x(t) - m_1(t)$ 满足 IMF 条件, 令 $c_1(t) = h_1(t)$, 那么 $c_1(t)$ 则为交通流量数据的 IMF1 分量。

重复以上步骤, 最后交通流量数据序列 $x(t)$ 可被分解成为:

$$x(t) = \sum_{i=1}^N c_i(t) + r_N(t) \quad (7)$$

3 EFLS - SVM 交通流量预测模型

3.1 LS - SVM 支持向量机

Suykens 提出的 LS - SVM 可转化为^[7-8]:

$$\begin{aligned} \min J(\omega, \xi) &= \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N \xi_k^2 \\ \text{s. t. } y_k &= \phi(x_k) \omega^T + b + \xi_k \end{aligned} \quad (8)$$

其中, $\xi_k \geq 0, k = 1, 2, \dots, N, C$ 为惩罚因子。

通过拉格朗日法可将式(8)转化成:

$$\begin{aligned} L(\omega, b, \xi, \alpha) &= \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N \xi_k^2 - \\ &\sum_{i=1}^N \alpha_k [(\omega^T \phi(x_k) + b + \xi_k) - y_k] \end{aligned} \quad (9)$$

其中, $\alpha_k (k = 1, 2, \dots, N)$ 表示拉格朗日乘子。

对式(9)进行 ω, b, ξ, α 求偏导, 并令其等于零, 则有:

$$\begin{aligned} \omega &= \sum_{k=1}^N \alpha_k \phi(x_k) \\ \sum_{k=1}^N \alpha_k &= 0 \\ \alpha_k &= C \xi_k \\ \omega^T \phi(x_k) + b + \xi_k - y_k &= 0 \end{aligned} \quad (10)$$

依据 Mercer 条件, $k(x_i, x_j)$ 核函数可以表示为:

$$k(x_i, x_j) = \phi(x_i) \phi(x_j) \quad (11)$$

本文运用 RBF 核函数实现预测, 表达式为:

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (12)$$

因此, LS - SVM 交通流量预测模型为:

$$f(x) = \sum_{i=1}^m \alpha_i \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) + \gamma \quad (13)$$

由式(13)可知,LS-SVM 的性能主要受 γ, σ 影响,为了实现 γ, σ 的自适应选择,本文运用 FOA 算法进行 γ, σ 进行自适应优化。

3.2 EFLS-SVM 网络预测模型

由于 LS-SVM 需要优化的参数为 γ, σ , 因此其优化的数学模型:

$$F = \{\gamma, \sigma\} \quad (14)$$

通过优化,在确保预测精度最优的情况下,实现 γ, σ 参数的自适应选择,其适应度函数可进行定义。假设 t 时刻的实际交通流量为 $y(t)$, 预测交通流量 $\hat{y}(t)$, 那么实际交通流量 $y(t)$ 和预测交通流量 $\hat{y}(t)$ 的差值 $e(t)$ 公式^[10]:

$$e(t) = \hat{y}(t) - y(t) \quad (15)$$

针对交通流量预测的非线性问题,实际交通流量数据样本为 n , 运用 FOA 优化 LS-SVM 的核参数和惩罚系数,使得 LS-SVM 的实际交通流量输出和预测交通流量之间的差值的平方和最小,适应度函数公式:

$$\min Fitness(t) = \frac{1}{2n} \sum_{t=1}^T [e(t)]^2 \quad (16)$$

3.3 算法步骤

基于 EFLS-SVM 的交通流量预测流程如图 1 所示,算法步骤如下:

Step1:归一化交通流量数据。

Step2:EMD 分解归一化的交通流量数据,提取交通流量数据的细节特征和趋势特征,构建出训练样本和测试样本。

Step3:设定 FOA 算法的最大迭代次数 $max\ gen$, 种群大小 $popsiz$ 。

Step 4:将构建出的训练样本输入 LS-SVM,根据适应度函数式(16)计算粒子的适应度函数值,寻找粒子个体和全局最优粒子的位置和最优值。

Step 5:粒子速度和位置的更新。

Step6:计算评估适应度大小并更新粒子的位置和速度。

Step7:若 $gen > max\ gen$, 保存最优解;反之 $gen = gen + 1$, 转到 Step4。

Step8:根据粒子的最优位置所对应的最优参数 γ, σ 实现交通流量的预测。

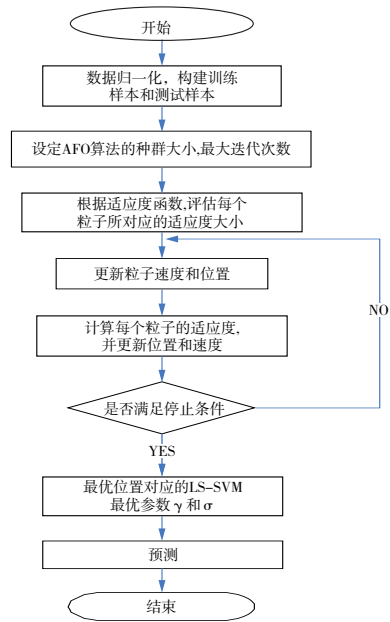


图 1 基于 EFLS-SVM 的预测流程图

4 仿真实验

4.1 数据来源

本文数据来源于某交通监控站监测数据,收集自 2015 年 9 月 7 日~21 日一共 15 天的数据为研究对象,每天每间隔 1 小时采集一次车流量数据,一共采集 $15 \times 24 = 360$ 组数据,交通流量的原始交通流量数据及其 EMD 分解情况如图 2 所示。

4.2 数据处理

实际交通流量数据序列进行 EMD 分解,依次可以分解出不同的 IMF 分量。由图 2(b)~图 2(f)可知,原始交通流量数据被分解成 4 个波动较小的分量和 1 个剩余分量。根据 IMF 分量的分析结果,运用 FOA 优化 LSSVM 的核参数和惩罚系数的模型进行交通流量预测。

4.3 评价指标

为了验证本文算法进行交通流量预测的有效性,本文采用均方误差来评价交通流量预测效果的评价指标,评价公式:

$$MSE = \sqrt{\frac{1}{K} \sum_{i=1}^K (x_i - \hat{x}_i)^2} \quad (17)$$

其中, x_i, \hat{x}_i 分别表示实际交通流量和预测交通流量。

4.4 实验结果

将收集的 360 组交通流量数据分成训练样本和测试样本,将前 336 组数据作为训练数据,用于建立预测

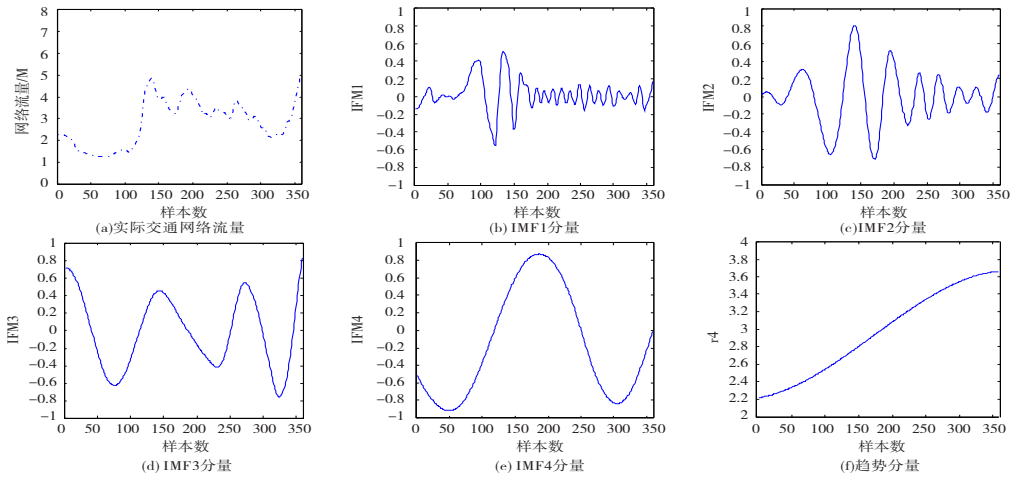


图2 实际交通流量和 EMD 分解结果图

模型;后面 24 组作为测试数据,用于验证预测结果的好坏。设定 FOA 算法的最大迭代次数为 100,种群大小为 20,其预测结果如图 3 ~ 图 6 所示,分别表示单步预测、3 步预测、5 预测和 7 步预测。

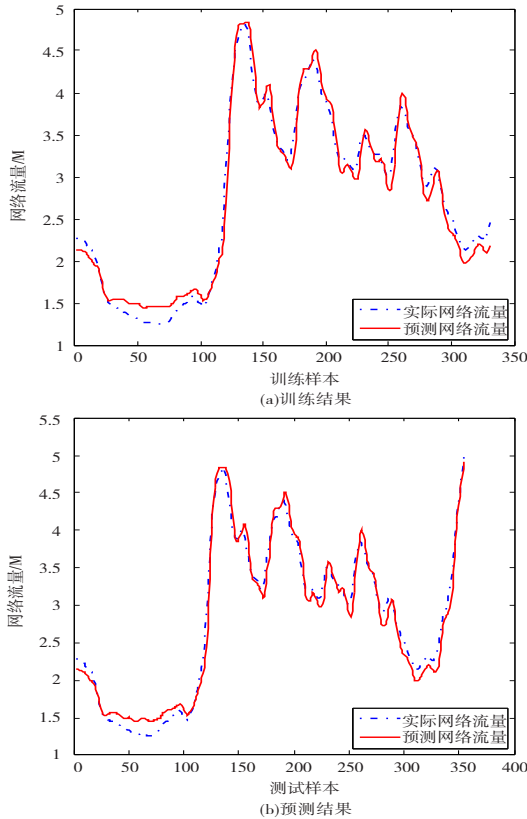


图3 基于 EFLS - SVM 算法的单步交通流量预测

由 EFLS - SVM 算法的单步预测、3 步预测、5 步预测和 7 步预测结果可知,随着预测步长的增加,EFLS - SVM 算法的预测精度不断提高,效果较好。图 7 表示 FOA 算法优化 LS - SVM 的适应度收敛曲线。

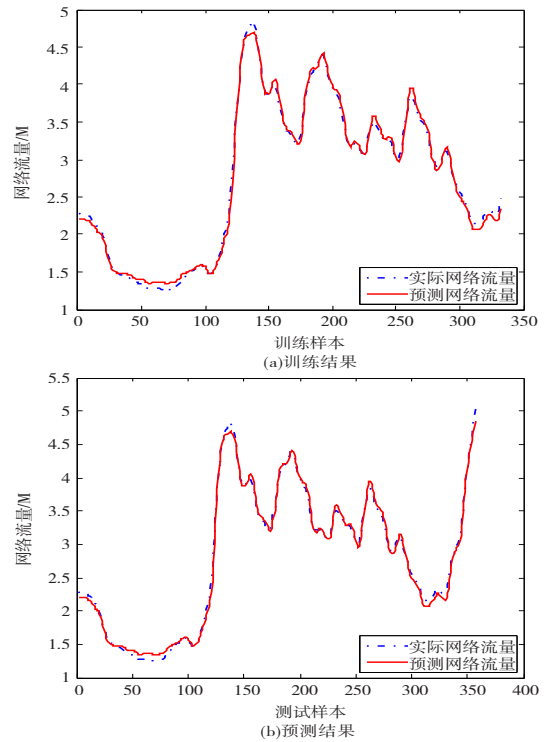


图4 基于 EFLS - SVM 算法的 3 步交通流量预测

为了验证 EFLS - SVM 算法的优越性,将 EFLS - SVM 算法、FOA - LSSVM 和 LS - SVM 三者的预测结果进行对比,运行 10 次,其对比结果见表 1。

由表 1 中 EFLS - SVM 算法、FOA - LSSVM 和 LSSVM 三者预测的 MSE 误差对比结果可知,EFLS - SVM 算法的预测效果最好,优于 FOA - LSSVM 和 LS - SVM 模型,其次 FOA - LSSVM 的预测效果优于 LS - SVM。

由表 2 中 EFLS - SVM 算法、FOA - LSSVM 和 LS - SVM 三种模型预测时间对比结果可知,EFLS - SVM 算法的预测时间最短快于 FOA - LSSVM 和 LS - SVM 模

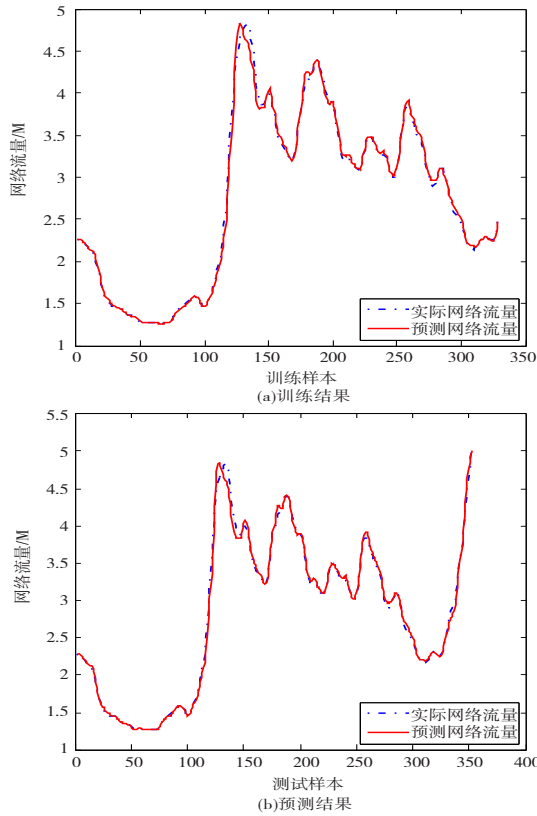


图 5 基于 EFLS-SVM 算法的 5 步交通流量预测

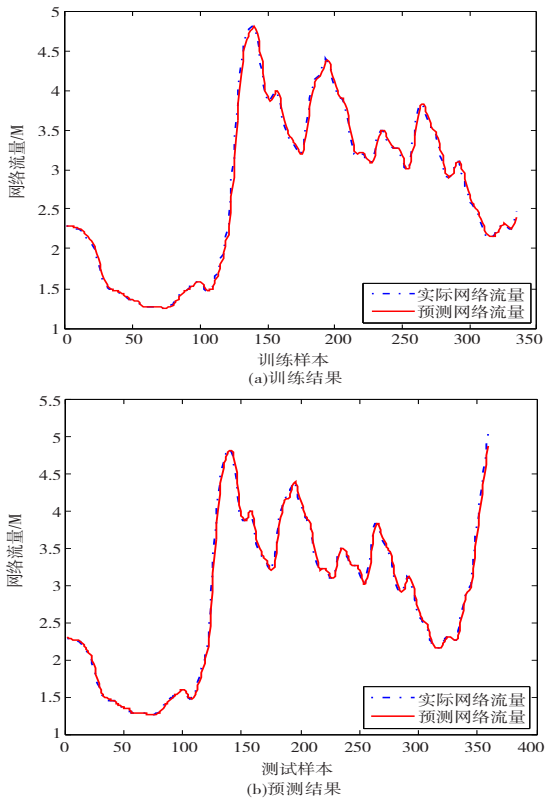


图 6 基于 EFLS-SVM 算法的 7 步交通流量预测

型,而 FOA-LSSVM 的预测时间短于 LSS-VM。

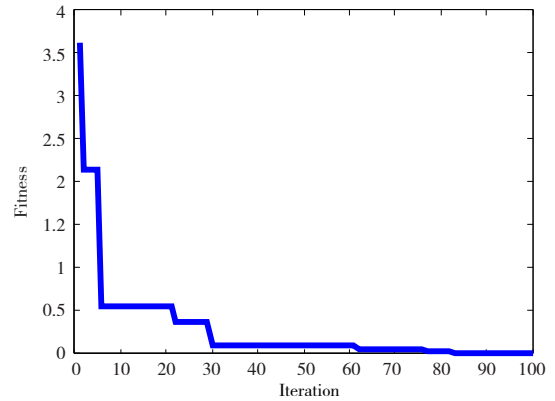


图 7 FOA 优化 LS-SVM 的适应度收敛曲线图

表 1 EFLS-SVM 算法、FOA-LSSVM 和 LS-SVM 三种模型预测 MSE 误差对比

运行次数	LS-SVM	FOA-LSSVM	EFLS-SVM
1	0.0088	0.0060	0.0040
2	0.0072	0.0055	0.0046
3	0.0068	0.0050	0.0040
4	0.0084	0.0064	0.0042
5	0.0065	0.0056	0.0038
6	0.0074	0.0061	0.0043
7	0.0076	0.0056	0.0035
8	0.0064	0.0046	0.0042
9	0.0072	0.0076	0.0040
10	0.0067	0.0051	0.0047
均值	0.0072	0.0058	0.0041

表 2 EFLS-SVM 算法、FOA-LSSVM 和 LSSVM 三种模型预测时间对比(单位:s)

预测步长	LS-SVM	FOA-LSSVM	EFLS-SVM
单步	111.40	97.36	37.21
3	97.60	89.22	34.25
5	86.33	74.22	32.18
7	79.45	65.80	31.27

为了进一步验证本文算法的有效性和可靠性,选择 CUDA 标准测试数据库为研究对象^[11-12],进行交通流量预测。将前 276 组数据作为训练数据,后面 92 组作为测试数据。设定 FOA 算法的最大迭代次数为 100,种群大小为 20,其预测结果如图 8 所示。

由图 8 可知,本文算法的预测精度较高,预测平均相对误差 6.71%,可以进行实际交通流量预测,不过存在个别点预测精度较低的缺点,可能受到其他影响因素的制约,如算法的性能、交通流量实际影响因素等。

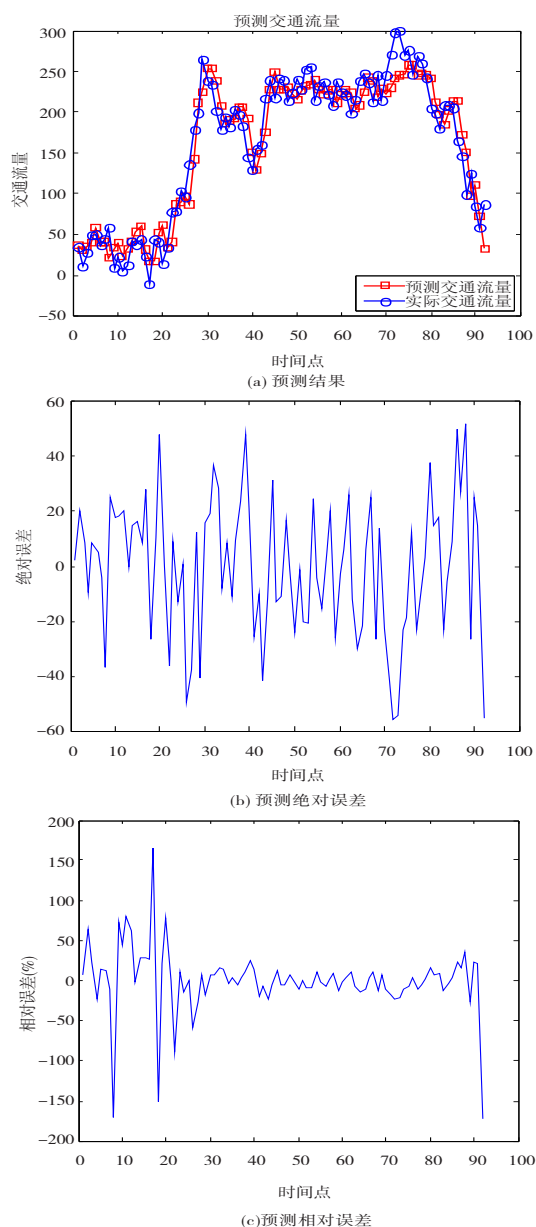


图8 CUDA标准数据集预测结果

5 结束语

针对LS-SVM核参数和惩罚系数选择的随机性,本文运用FOA算法对LS-SVM核参数和惩罚系数进行优化,同时结合EMD提取交通网络流量的细节特征和趋势特征,构建出基于EFLS-SVM的交通网络流量预测模型,分别进行单步、3步、5步和7步预测。通过对不同交通网络流量预测模型预测均方误差和预测时间发现,EFLS-SVM算法的预测精度和预测效率均优于其他模型,从而为交通网络资源的合理配置提供科学决策的依据。

参考文献:

- [1] 雷霆,余镇危.一种交通流量预测的小波神经网络模型[J].计算机应用,2012,26(3):526-528.
- [2] Guo Wen,Qiao Yizheng,Hou Haiyan.BP neural network optimized with PSO algorithm and its application in forecasting[C]//Proceedings of the IEEE international Conference on Information Acquisition,Weihai,China,January 1,2011:617-621.
- [3] 刘渊,戴悦,曹建华.基于小波神经网络的流量混沌时间序列预测[J].计算机工程,2012,34(16):105-110.
- [4] 杨光,张国梅,刘星宇.基于小波核LS-SVM的交通流量预测[J].微机发展,2011,15(12):125-128.
- [5] Pan W T.A new fruit fly optimization algorithm:Taking the financial distress model as an example[J].Knowledge-Based Systems,2012,26:69-74.
- [6] 王西锋,高岭,张晓亭.自相似交通网络流量预测的分析和研究[J].计算机技术与发展,2010,17(11):42-45.
- [7] 冯海亮,林青家,陈涤,等.一种基于神经网络的交通网络流量组合预测模型[J].计算机应用,2010,26(9):2206-2208.
- [8] Cervellera C,Wen A H,Chen V C P.Neural network and regression spline value function approximations for stochastic dynamic programming[J].Computer & Operations Research,2012,34(1):70-90.
- [9] Li Junfeng,Yang Aiping,Dai Wenzhan.Modeling mechanism of Grey Neural Network and its application[C]//Proceedings of 2007 IEEE International Conference on Grey Systems and Intelligent Services, Nanjing, November 18-20,2009:404-408.
- [10] 李宗福,邓琼波,李桓.Kohonen SOFM神经网络及其演化研究[J].计算机工程与设计,2013,25(10):1729-1830.
- [11] Wang H,Liu Y,Zeng S Y.A Hybrid Particle Swarm Algorithm with Cauchy Mutation[C]//Proceeding of

- IEEE Swarm Intelligence Symposium (SIS), Honolulu, HI, April 1-5, 2011: 356-360.
- [12] Wang H, Liu Y, Wu Z, et al. An improved particle swarm optimization with adaptive jumps[C]//Proceedings of IEEE International Conference on Evolutionary Computation, Brisbane, Australia, June 10-15, 2012: 392-397.

Predictive Analysis of Traffic Flow Based on Improved LS-SVM Mathematical Model

WU Yifan

(Department of Basic Courses, Jiangsu Food & Pharmaceutical Science College, Huai'an 223003, China)

Abstract: According to the traffic flow data having nonlinear and non-stationary characteristics, the EMD and FOA algorithm were used to implement self-adaptive optimization selection of LS-SVM kernel parameters and penalty coefficient, and then a traffic flow forecasting model based on EFLS-SVM algorithm was proposed. The minutiae characteristics and trend feature of traffic flow data were extracted by EMD, and a traffic prediction model based on EFLS-SVM was built, then single-step, three-step, five-step and seven-step prediction were proceeded respectively. By comparing different experiments, the results showed that prediction accuracy and prediction efficiency of EFLS-SVM algorithm were better than other models, thus scientific decision-making basis was provided for the rational allocation of transport network resources.

Key words: traffic flow; fruit fly optimization algorithm; mathematical model; least squares support vector machine