

# SVR Based on FOA and Its Application in Traffic Flow Predication

Wei Zhu, Nan Li, Chaofeng Shi, Bingfeng Chen

School of Traffic & Transportation, Chongqing Jiaotong University, Chongqing  
Email: zrcqxnu@163.com

Received: Dec. 2<sup>nd</sup>, 2012; revised: Jan. 5<sup>th</sup>, 2013; accepted: Jan. 15<sup>th</sup>, 2013

**Abstract:** The accurate predication of traffic flow is an essential job in ITS. In order to predict traffic flow exactly, the FOA-SVR prediction model combining Support Vector Regression (SVR) and fruit fly optimization algorithm (FOA) is presented to forecast railway traffic flow. Using fruit fly optimization algorithm to optimize training parameters of Support Vector Regression, which can obtain superior SVR prediction model. The experiment results show that the FOA-SVR model has more accuracy, which provides a new approach for traffic flow prediction.

**Keywords:** Fruit Fly Optimization Algorithm; Traffic Flow Forecasting; Support Vector Regression (SVR); Optimization Parameters

## FOA-SVR 在交通流预测中的研究

朱 伟, 李 楠, 石超峰, 陈丙锋

重庆交通大学交通运输学院, 重庆  
Email: zrcqxnu@163.com

收稿日期: 2012 年 12 月 2 日; 修回日期: 2013 年 1 月 5 日; 录用日期: 2013 年 1 月 15 日

**摘 要:** 交通流量预测是实现智能交通系统的重要工作。为了更准确地对交通流量进行预测, 结合支持向量回归机(SVR)与果蝇算法(FOA), 提出了 FOA-SVR 的交通流量预测模型。利用果蝇算法优化支持向量回归机的训练参数, 以得到预测性能更好的支持向量回归预测模型。仿真结果表明, FOA-SVR 模型比传统的支持向量机模型预测精度更高, 可以更好的对交通流进行预测。

**关键词:** 果蝇算法; 交通流预测; 支持向量回归机; 优化参数

### 1. 引言

短时交通流预测一直是国内外的研究热点。由于交通流的产生有很大的随机性, 离散性, 非线性, 常见的预测方法, 如 ARMA 方法<sup>[1]</sup>和卡尔曼滤波法<sup>[2]</sup>等预测结果难以满足交通流预测的精度要求。随着智能理论和计算机网络的发展, 神经网络开始被广泛地应用于交通流预测中, 它具有良好的非线性映射能力, 但是也存在着收敛速度慢, 过学习和局部极值等问题<sup>[3]</sup>, 这些问题极大影响了神经网络的预测精度。

支持向量机(Support Vector Machine, SVM)是基于统计学习理论建立起来的机器学习算法, 不存在局部极值问题, 计算复杂性与输入样本维数无关, 具有很强的泛化能力和抗噪声能力, 是处理小样本数据学习问题的有力工具, 在文本分类、图像识别、函数估计、非线性建模等领域获得广泛的应用<sup>[4]</sup>。但是和其它算法一样, 支持向量机的性能依赖于学习机的参数, 即参数的选择对支持向量机的预测效果有着较大的影响。因此, 根据实际的数据模型选择合适的训练

参数成为支持向量机预测关键问题<sup>[5]</sup>。本文在分析支持向量回归机(SVR)理论上,提出了用果蝇算法 FOA 优化参数的 FOA-SVM 交通流预测模型,使得 SVR 的参数选取由人工选取变为自动确定,用实际数据仿真表明了该模型的有效性。

## 2. 支持向量回归机的基本原理

SVM 本身是针对经典的二分类问题提出的,支持向量回归机(Support Vector Regression, SVR)是支持向量在函数回归领域的应用。SVR 与 SVM 分类有以下不同: SVM 回归的样本点只有一类,所寻求的最优超平面不是使两类样本点分得“最开”,而是使所有样本点离超平面的“总偏差”最小,此时样本点都在两条边界线之间,求最优回归超平面同样等价于求最大间隔。本文采用的支持向量机在回归分析中的一种算法  $\varepsilon$ -支持向量机( $\varepsilon$ -SVR)。

$\varepsilon$ -SVR 的算法原理如下: 设训练集

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (R^n \times Y)^l \quad (1)$$

其中,  $x_i \in R^n, y_i \in Y, i=1, \dots, l$ , 对于线性回归问题, 主要是在  $R^n$  上寻找一个线性函数  $g(x) = (\omega x) + b$ , 以便用  $y = g(x)$  来计算任一输入  $x$  所对应的输出值  $y$ 。对于非线性问题, 我们则需要引进一个变换  $\varphi(x)$  将训练数据映射到一个高维空间, 把原空间的非线性回归转换成高维空间的线性回归<sup>[4]</sup>。这一变换主要通过核函数来实现。目前, 常用的核函数主要有以下几种:

- 1) 多项式核函数  $K(x, x') = [(x, x') + l]^d$
- 2) 径向基(RBF)函数

$$K(x, x) = \exp\left\{-\frac{1}{\sigma^2} \|x - x_i\|^2\right\}$$

- 3) Sigmoid 核函数

$$K(x, x) = \tanh(v(x, x_i) + c)$$

选取适当的核函数  $K(x, x)$  以及适当的精度  $\varepsilon > 0$  和惩罚参数  $C > 0$ , 构造并求解凸二次规划问题<sup>[5]</sup>。

$$\min_{\alpha^{(*)} \in R^{2l}} \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) K(x_i, x_j) + \varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l y_i (\alpha_i^* - \alpha_i) \quad (2)$$

$$s.t. \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad 0 \leq \alpha_i^* \leq C, i=1, 2, \dots, l \quad (3)$$

得解  $\bar{\alpha}^{(*)} = (\bar{\alpha}_1, \bar{\alpha}_1^{(*)}, \dots, \bar{\alpha}_l, \bar{\alpha}_l^{(*)})$ 。计算  $\bar{b}$ , 选取位于区间  $(0, C)$  中  $\bar{\alpha}^{(*)}$  的分量  $\bar{\alpha}_j$  或  $\bar{\alpha}_k^*$ , 如果选到的是  $\bar{\alpha}_j$ , 则

$$\bar{b} = y_j - \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) K(x_i, x_j) + \varepsilon \quad (4)$$

如果选到的是  $\bar{\alpha}_j$ , 则

$$\bar{b} = y_k - \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) K(x_i, x_k) - \varepsilon \quad (5)$$

最后构造出决策函数:

$$y = g(x) = \sum_{i=1}^l (\bar{\alpha}_i^* - \bar{\alpha}_i) K(x_i, x) + \bar{b} \quad (6)$$

损失函数采用  $\varepsilon$ -不敏感损失函数, 对于  $\varepsilon$ -支持向量回归机, 其预测精度受参数影响较大, 因此需要选择合适的核函数类型及其参数组合。本文采用径向基函数(RBF)为核函数, 所以需要确定的参数有惩罚系数  $C$  和不敏感损失函数宽度  $\varepsilon$ 。由于传统的参数选取基本都是根据交叉验证法来选取参数, 但是这种方法选取的时间较长并且无法保证取得最优参数值。为了提高参数的选择效率, 需要采用合适的智能优化算法搜寻支持向量回归机参数的最优组合, 从而获得具有较好预测能力的预测模型。

## 3. 基于果蝇优化算法的 SVR 参数优化模型

### 3.1. 果蝇优化算法

果蝇优化算法(Fruit Fly Optimization Algorithm, FOA)是台湾学者潘文超提出的, 它是一种基于果蝇觅食行为推演出的寻求全局最优化的新方法<sup>[6]</sup>。由于果蝇本身在感观知觉上优于其它物种, 尤其在视觉与嗅觉上, 果蝇的嗅觉器官能很好地搜集漂浮在空气中的各种气味, 通过使用其灵敏的视觉发现食物与同伴的聚集位置, 最后找到味道浓度最高的果蝇<sup>[7]</sup>。

### 3.2. 果蝇算法优化 SVR 参数步骤

本文采用 FOA 算法来对 SVR 的参数对  $(C, \varepsilon)$  进行优化。在 FOA-SVR 算法中, 每一个果蝇味道浓度判定值代表 SVR 的一组参数对  $(C, \varepsilon)$ 。将 FOA 算法

应用于 SVR 参数优选时, 算法步骤如下:

步骤 1, 读入样本集。

步骤 2, 确定种群个体数量  $sizepop = 20$  和最大迭代次数  $max\ gen = 100$ , 在  $[0,1]$  范围内, 随机生成果蝇的初始位置。

步骤 3, 赋予果蝇个体搜寻食物的随机飞行方向与距离区间。

步骤 4, 估计果蝇位置与原点之间的距离, 计算味道浓度判定值( $S_i$ )。  $S_i = \frac{1}{Dist_i}$ 。

步骤 5, 将参数组合  $(C, \epsilon)$  代入 SVR 预测模型中, 以预测的误差平方和作为味道判定函数, 求出该果蝇位置味道浓度( $Smell(i)$ ), 即误差平方和。

步骤 6, 找出果蝇群体使得误差平方和最小的果蝇, 即其味道浓度最低。

步骤 7, 保留最佳模型参数  $(C, \epsilon)$  与  $(X_i, Y_i)$  坐标, 此时果蝇群体利用视觉往该位置飞去

步骤 8, 迭代寻优, 重复执行步骤 3~6, 并判断预测误差平方和是否优于前一迭代预测误差平方和, 若是则执行步骤 7。

#### 4. FOA-SVR 的短时交通流量预测

交通流量预测就是建立从嵌入  $m$  维相空间到一维的回归模型<sup>[8]</sup>。下面以文献[9]的第 23 章小波神经网络的时间序列预测—短时交通量预测。以 15 min 为时间间隔采集 4 天共 384 个数据, 用前三天的数据训练神经网络, 最后用第 4 天的数据来验证模型的准确度。根据交通量的周期特性设计 FOASVR 预测模型。输入层为当前时间点前 4 个时间点交通量数据, 输出层为当前时间点的预测交通量。

##### 1) 数据准备和预处理

为了加快网络的收敛速度, 需要对输入输出数据进行归一化处理。本文对样本的输入和输出数据在区间  $[0,1]$  进行归一化处理。

##### 2) 利用 FOA 算法进行参数寻优

利用 FOA 算法对 SVR 的训练参数进行优化, 种群规模  $Sizepop = 20$ , 迭代次数  $Maxgen = 100$ , 初始果蝇飞行方向与距离区间为  $[0,1]$ 。

应用训练好的模型对后第 4 天的测试样本进行检验, 预测结果如图 1 所示。

从图 1 预测结果可以看出, 基于果蝇优化 SVR

的交通流量预测方法能够准确地预测短时交通流量, 预测值与期望值比较接近, 说明了用 FOA 优化 SVR 的短时交通流预测模型具有一定的可行性。

为了更进一步验证 FOASVR 的优越性, 我们同时用文献[9]的小波神经网络预测进行比较, 误差结果的对比指标采用了平均绝对偏差(Mean Absolute Deviation, MAD, 表示预测值与实际值的偏差绝对值的均值)和平均绝对百分比误差(Mean Absolute Percentage Error, MAPE, 表示预测值与实际值的偏差占实测值的百分比的均值), 其结果见表 1。

对比这 2 种方法的预测结果, 可以看出 FOASVR 预测的每项指标均比小波神经网络有更好的预测效果。

#### 5. 研究结果

交通流量具有一定的非线性、周期性和随机性, 传统的预测方法难以发现它们之间的规律, 本文利用果蝇算法和 SVR 相结合构建 FOASVR 预测的新方法, 进行短时交通流预测的研究。果蝇算法能够快速找出 SVR 的最优训练参数, 从而使预测的结果更加精确。结果表明, 相比于小波神经网络, FOASVR 的预测方

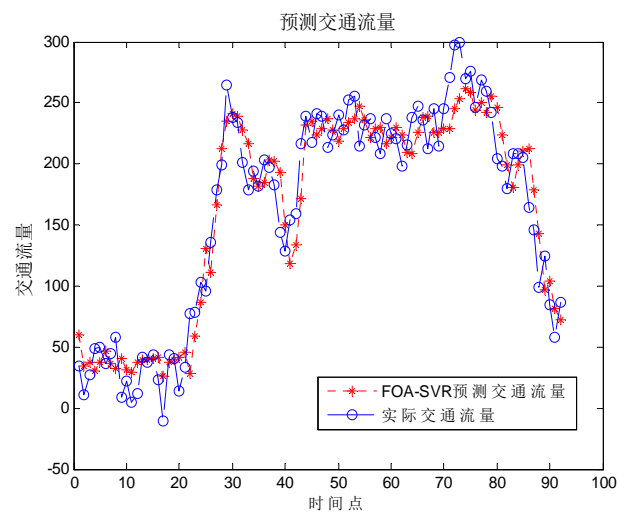


Figure 1. The prediction result of FOA-SVR  
图 1. FOA-SVR 预测结果图

Table 1. The error index comparison of two kinds of models  
表 1. 两种预测方法误差指标对比

算法	MAPE	MAD
小波神经网络	17.58%	20.551
FOASVR	15.6%	19.563

法预测精度更高,说明该模型是一种有效,可行的方法,在短时交通量预测中具有一定的推广价值。

### 参考文献 (References)

- [1] 贺国光,李宇,马寿峰.基于数学模型的短时交通流预测方法探讨[J].系统工程理论与实践,2000,12:51-56.
- [2] 杨兆升,朱中.基于卡尔曼滤波理论的交通流量实时预测模型[J].中国公路学报,1999,12(3):63-67.
- [3] 文培娜,张志勇,罗彬.基于BP神经网络的北京物流需求预测及分析[J].物流技术,2009,6:91-93.
- [4] V. Vapnik. An overview of statistical learning theory. IEEE Transactions on Neural Networks, 1999, 10(5): 988-999.
- [5] 孙煦,陆化普,吴娟.基于蚁群优化算法的公路客运量预测[J].合肥工业大学学报,2012,1:124-129.
- [6] W.-T. Pan. A new fruit fly optimization algorithm: Taking the financial distress model as an example. Knowledge-Based Systems, 2011, 26: 69-74.
- [7] 许智慧,王福林,孙丹丹,王吉权.基于FOA-RBF神经网络的外贸出口预测[J].数学的实践与认识,2012,13:14-19.
- [8] 张锐,张涛,高辉.RQEA-SVR在交通流预测中的作用[J].计算机工程与应用,2010,46(9):241-245.
- [9] 史峰,王小川,郁磊,李洋. Matlab 神经网络 30 个案例分析 [M]. 北京:北京航空航天大学出版社,2011.