

# 改进支持向量机的网络流量预测<sup>①</sup>

王雪松

(佛山职业技术学院 电子信息系, 佛山 528137)

**摘要:** 支持向量机具有良好的非线性建模能力, 其参数对网络流量预测结果有直接影响, 为了解决支持向量机的参数确定的难问题, 根据杂草算法的优势, 提出了改进支持向量机的网络流量预测模型. 首先收集大量网络数量原始数据, 将支持向量机参数作为杂草种子, 然后模拟杂草的生存、繁殖过程搜索最优参数寻优, 建立网络流量预测模型, 最后采用具体网络流量数据测试模型的可行性. 结果表明, 该模型不仅得到了高精度的网络流量预测结果, 而且可以应用网络流量管理中.

**关键词:** 网络流量; 杂草算法; 混沌理论; 支持向量机

## Network Traffic Predicting Model Based on Improved Support Vector Machine

WANG Xue-Song

(Department of Electronic Information, Foshan Polytechnic College, Foshan 528137, China)

**Abstract:** Aiming at parameters optimization problem of support vector machine in network traffic predicting, a network traffic predicting model is proposed based on improved support vector machine. Parameters of support vector machine are considered as a weed, the optimal parameters are found by invasive weed optimization algorithm, and network traffic data is used to test the performance. The experimental results show that the proposed model obtains high predicting accuracy and fastens the model speed, and it can meet the requirements of network traffic predicting.

**Key words:** network traffic; invasive weed optimization algorithm; chaotic theory; support vector machine

随着 Internet 的数据不断增加, 网络管理面临巨大的挑战, 为了防止出现网络拥堵现象, 网络流量预测引起了人们的高度关注. 流量预测从历史数据中发现将来网络状态的变化趋势, 为企业和管理人员提供有意义的参考信息, 已经成为网络管理领域中的研究重点<sup>[1,2]</sup>.

当前网络流量预测模型分为两类: 传统模型和现代, 传统模型主要有: 多元线性回归算法、时间序列分析法等<sup>[3,4]</sup>, 根据历史数据确定网络流量预测模型的参数, 建模速度快、硬件要求低, 然而它们均基于网络流量是一种固定变化趋势, 如周期性、单调递增等, 不能描述网络流量随机变化特点, 预测精度低<sup>[5]</sup>. 现代模型采用现代统计学对网络流量进行建模与分析, 主要有: 神经网络、支持向量机(support vector machine,

SVM)等, 可以反映网络流量变化的随机性、时变性, 网络流量预测结果更加靠, 获得了较高的预测精度<sup>[6-8]</sup>. SVM 具体要求历史样本少, 泛化能力优势, 成为网络流量当前主要的建模工具, 然而网络流量预测精度与核函数及相关参数密切相关, 因此 SVM 参数确定是网络流量建模和预测首先要解决的问题<sup>[9]</sup>. 为了解决 SVM 参数确定问题, 业内学者们采用进化算法、粒子群算法等实现<sup>[10]</sup>, 它们得到了较好的 SVM 参数, 但它们同样存在不足, 如进化算法的收敛速度慢, 遗传算法参数确定没有理论指导, 凭经验随机确定, 通用性差, 难以找到最优的 SVM 参数, 影响网络流量的预测结果.

为了获得更优的网络流量预测结果, 针对 SVM 参数优化难题, 提出了改进支持向量机的网络流量预测

① 基金项目:广东省教育厅项目(2010TJK446)

收稿时间:2016-06-09;收到修改稿时间:2016-08-08 [doi: 10.15888/j.cnki.csa.005668]

模型(IWO-SVM),采用杂草优化算法(invasive weed optimization algorithm, IWO)搜索最优的 SVM 参数,采用网络流量预测实验测试可行性和可靠性。

## 1 支持向量机和杂草优化算法

### 1.1 支持向量机(SVM)

设训练样本集为:  $\{(x_i, y_i)\}$ ,  $i=1, 2, \dots, n$ , 根据函数  $\phi(x)$  将其映射到高维特征空间实现回归, 即:

$$f(x) = w^T \phi(x) + b \quad (1)$$

式中,  $w$  为权值向量;  $b$  为偏置向量<sup>[11]</sup>。

将式(1)转换为最小化问题, SVM 回归的目标函数为:

$$\begin{aligned} \min & \|w\|^2 + \frac{1}{2} C \sum_{i=1}^n \xi_i^2 \\ \text{s.t.} & \\ & \begin{cases} y_i - w^T \phi(x) + b = e_i \\ i = 1, 2, \dots, n \end{cases} \end{aligned} \quad (2)$$

式中,  $C$  为惩罚参数;  $e_i$  为回归误差。

采用拉格朗日算子( $\alpha_i$ )进行变换, 得到相应对偶问题为:

$$\begin{aligned} L(w, b, \zeta, \alpha) = \min & \|w\|^2 + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2 + \\ & \sum_{i=1}^n \alpha_i (\omega^T \phi(x) - b + e_i - y_i) \end{aligned} \quad (3)$$

设核函数  $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ , 采用 RBF 函数建立非线性的 SVM 回归模型, 则有:

$$f(x) = \sum_{i=1}^N \alpha_i \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) + b \quad (4)$$

式中,  $\sigma$  为核宽度。

对 SVM 的工件原理进行分析可知, 预测结果与参数  $C$  和  $\sigma$  相关, 为了克服当前算法存在的不足, 利用杂草算法搜索能力强、速度快的优点, 解决参数  $C$  和  $\sigma$  选择问题, 建立更加科学的网络流量预测模型。

### 1.2 杂草优化算法(IWO)

受到杂草生长和繁殖过程启发, 有学者提出了杂草优化算法<sup>[12]</sup>, 具体过程如下:

1) 每一个杂草( $x_i$ )在领域内繁殖一些种子, 种子数量  $\omega_i$  计算公式为

$$\omega_i = \frac{f(x_i) - f_{\min}}{f_{\max} - f_{\min}} (s_{\max} - s_{\min}) + s_{\min} \quad (5)$$

式中,  $f(x_i)$  为杂草  $x_i$  的适应度值;  $s_{\min}$  和  $s_{\max}$  分别为杂草

最小和大种子数;  $f_{\min}$  和  $f_{\max}$  为种群的最小和最大的适应度值。

2) 种子服从  $N(0, \sigma)$  分布, 其中标准偏差  $\sigma$  的计算公式为:

$$\sigma_{\text{iter}} = \frac{(\text{iter}_{\max} - \text{iter})^n}{(\text{iter}_{\max})^n} (\sigma_{\text{initial}} - \sigma_{\text{final}}) + \sigma_{\text{final}} \quad (6)$$

式中,  $\sigma_{\text{initial}}$  和  $\sigma_{\text{final}}$  为  $\sigma$  的初值和终值;  $\text{iter}_{\max}$  和  $\text{iter}$  为最大进化代数 and 当前进化代数;  $n$  为调和参数。

3) 当种群的种子数超过最大数量  $P\_Max$ , 对杂草和种子进行排序, 选择前  $P\_Max$  个适应度值最大的个体产生新的种群, 丢弃其它个体。

## 2 IWO-SVM的网络流量预测

由于参数  $C$  和  $\sigma$  的优化目标是提高流量预测的精度最高, 因此  $C$  和  $\sigma$  优化数学模型可以描述为:

$$\begin{aligned} \min & f(C, \delta) = \sum_{i=1}^n (y_i - y_i')^2 \\ \text{s.t.} & \end{aligned} \quad (7)$$

$$\begin{cases} C \in [C_{\min}, C_{\max}] \\ \delta \in [\delta_{\min}, \delta_{\max}] \end{cases}$$

式中,  $y_i$  和  $y_i'$  分别表示实验值和估计值。

基于 IWO-SVM 的网络流量预测步骤为:

1) 收集某一段时间的网络流量历史数据, 为了防止数据值间的差异过大进行预处理, 具体为:

$$\hat{y}_i = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}} \quad (8)$$

式中,  $\hat{y}_i$  为归一化值;  $y_{\min}$  和  $y_{\max}$  为原始数据的最小和最大值。

2) 网络流量受到人为因素、上网时间等影响, 具有一定的混沌性能, 因此需要对原始网络流量时间序列进行相空间重构<sup>[13]</sup>, 产生支持向量机的训练和验证样本。

3) 初始化杂草种群, 并初始化其它参数的值, 个体与一组  $(C, \sigma)$  相对应。

4) 支持向量机根据每一组  $(C, \sigma)$  对训练样本集进行建模, 计算网络流量的预测精度, 得到相应的适应度值。

5) 判断是否达到结束条件, 若达到就输出最优个体, 否则执行步骤(6)。

6) 产生新的种子, 并与其它种子和杂草组合, 选择  $P_{\max}$  个优秀个体组成新的种群, 返回步骤(4)。

7) 最优个体对应的  $C$  和  $\sigma$  为支持向量机的最优参数, 根据对训练样本进行学习建立网络流量预测模型.

### 3 网络流量预测模型性能的测试

#### 3.1 源数据

选择一台网络服务器出口的每小时网络流量进行仿真实验, 得到 1000 个样本, 具体如图 1 所示. 选择前 800 个数据点作为训练样本集构建网络流量预测模型, 其它数据点作为验证集分析模型的性能, 在 VC++ 6.0 平台实现仿真实验.

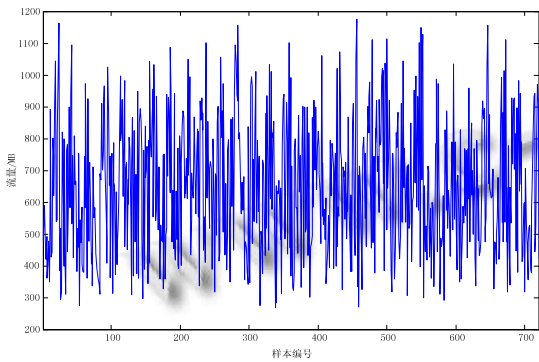


图 1 网络流量数据

#### 3.2 训练样本和验证样本的重构

通常情况下, 网络流量有弱混沌特性, 设延迟时间  $\tau=1$ , 采用假近邻算法估计最优嵌入维数( $m$ ), 结果如图 2 所示, 从图 2 可以清楚看出最优  $m=5$ , 根据  $\tau=1$ ,  $m=5$  对原始网络流量数据进行相空间重构, 得到重构后的训练样本和验证样本, 将其变为有规律的数据, 便于支持向量机进行建模.

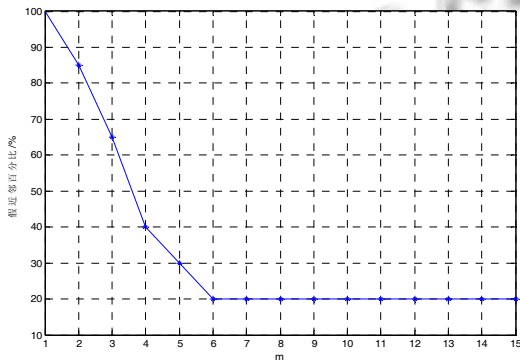


图 2 估计最优  $m$

#### 3.3 IWO 估计最优( $C, \sigma$ )SVM 参数

采用 IWO 算法搜索支持向量机的参数  $C$  和  $\sigma$  的最

优值, 1 步和 3 步预测的最优  $C$  和  $\sigma$  如表 1 所示.

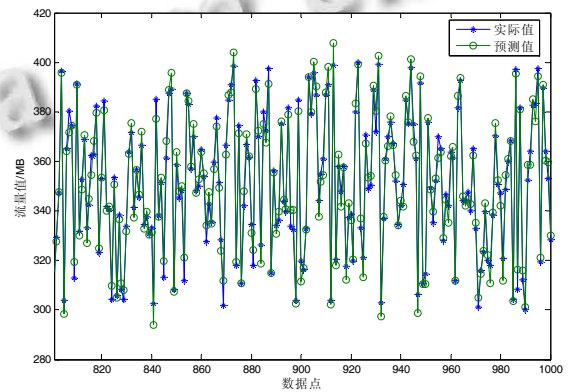
表 1  $C$  和  $\sigma$  的最佳值

预测步长	$C$	$\Sigma$
1	100.25	1.29
3	128.70	0.93

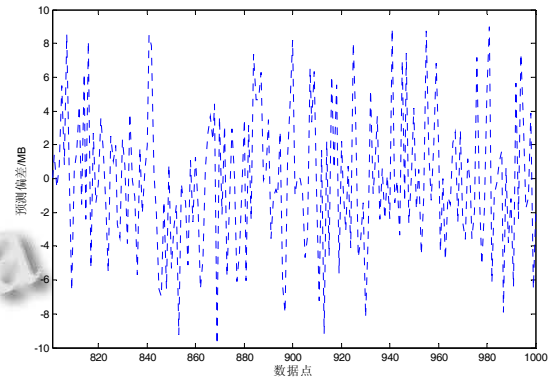
#### 3.4 结果与讨论

##### 3.4.1 本文模型的预测结果

IWO-SVM 的网络流量单步预测结果如图 3 所示, 单步预测值与实际值变化趋势相似, 二者间的误差小, 可以反映网络流量数据的时变性, 预测结果可信.



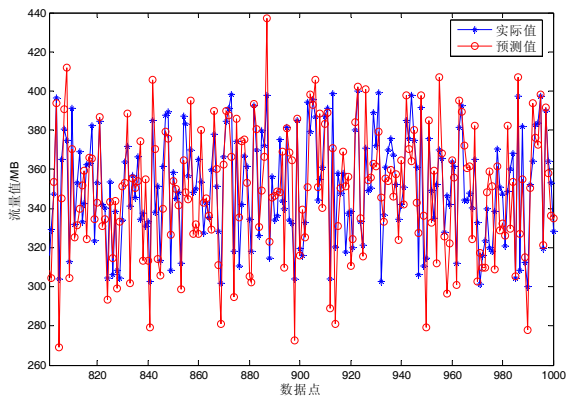
(a) 预测结果



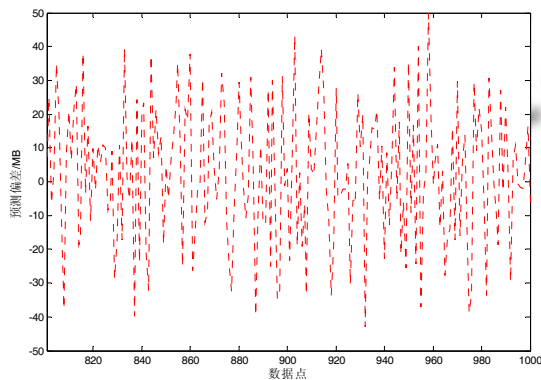
(b) 预测偏差

图 3 IWO-SVM 的网络流量 1 步预测结果

1 步预测只能提前一个时刻对网络流量变化趋势进行描述, 不能满足网络管理的要求, 因此实现了提前 3 步的网络流量预测实验, 结果如图 4 所示, 3 步预测精度明显要小于 1 步预测精度, 预测偏差也增大, 然而 IWO-SVM 还是能够描述网络流量整体变化趋势, 预测结果可以为网络管理者提供有用的信息.



(a) 预测结果



(b) 预测偏差

图 4 IWO-SVM 的网络流量 3 步预测结果

### 3.4.2 与经典模型的性能对比

选择当前经典网络流量预测模型进行对照实验, 它们为: ARIMA、BP 神经网络, 采用均方根误差 (RMSE) 和相对平均误差 (MPAE) 作为性能评价指标, 结果见表 2. IWO-SVM 的 MPAE 和 RMSE 均要小于经典模型的 MPAE 和 RMSE, 网络流量预测精度更高, 说明 IWO-SVM 是一种精度高、结果可靠的网络流量预测模型.

表 2 网络流是的 MPAE 和 RMSE 对比

预测模型	1 步		3 步	
	RMSE	MAPE/%	RMSE	MAPE/%
ARIMA	22.49	5.80	50.21	10.37
BP 神经网络	17.36	3.74	44.25	8.24
IWO-SVM	4.12	0.97	20.88	4.92

## 4 结语

针对网络流量的 SVM 参数优化问题, 提出了基于 IWO-SVM 的网络流量预测模型, 采用杂草算法搜索 SVM 参数, 解决当前参数选择的盲目性, 实验结果表明, IWO-SVM 的网络流量预测精度高, 预测结果好于经典模型, 具有广泛的应用前景.

### 参考文献

- 史振华, 刘外喜, 杨家焯. SDN 架构下基于 ICMP 流量的网络异常检测方法. 计算机系统应用, 2016, 25(4): 135-142.
- 郭平, 吴斌. 采用回归方法优化网络流量管理模型处理性能. 计算机工程与应用, 2012, 48(4): 104-106.
- 姜明, 吴春明, 张曼, 胡大民. 网络流量预测中的时间序列模型比较研究. 电子学报, 2009, 37(11): 2353-2358.
- 陈晓大, 刘静娴. 改进的基于小波变换和 FARIMA 模型的网络流量预测算法. 通信学报, 2011, 32(4): 153-157.
- 麻书钦, 范海峰. 基于小波变换和时间序列的网络流量预测模型. 河南理工大学学报: 自然科学版, 2013, 32(2): 188-192.
- 曲桦, 马文涛, 赵季红, 等. 基于最大相关熵准则的网络流量预测. 高新技术通讯, 2013, 23(1): 134-145.
- 王雪松, 梁昔明. 基于 BPSO-RBF 神经网络的网络流量预测. 计算机应用与软件, 2014, 31(9): 102-105.
- 肖汉杰, 桑秀丽. 相关向量机超参数优化的小时间尺度网络流量非线性预测方法. 计算机应用研究, 2016, 33(6): 1882-1885.
- 于明, 艾月乔. 基于人工蜂群算法的支持向量机参数优化及应用. 光电子激光, 2012, 23(2): 374-398.
- 邵信光, 杨慧中, 陈刚. 基于粒子群优化算法的支持向量机参数选择及其应用. 控制理论与应用, 2006, 23(5): 740-748.
- 刘忠宝. 新型支持向量机在风速预测模型中的应用研究. 电子科技大学学报, 2014, 43(5): 754-758.
- Mallahzadeh AR, Oraizi H, Davoodi RZ. Application of the invasive weed optimization technique for antenna configurations. Progress in Electromagnetic Research, 2008, 20(79): 137-150.
- 黄发明, 殷坤龙, 张桂荣, 等. 基于相空间重构和小波分析-粒子群向量机的滑坡地下水位预测. 地球科学-中国地质大学学报, 2015, 40(7): 1254-1265.