

# 基于 PSO-LSSVM 的网络流量预测模型<sup>①</sup>

刘 春

(四川建筑职业技术学院 网络管理中心, 德阳 618000)

**摘 要:** 为了提高网络流量的预测精度, 考虑到网络流量的长相关、非线性等特性, 提出一种粒子群算法优化最小二乘支持向量机参数的网络流量预测模型(PSO-LSSVM)。首先将最小二乘支持向量机参数作为粒子的位置向量, 然后利用粒子群算法找到模型的最优参数, 最后采用最优参数最小二乘支持向量机建立网络流量预测模型。仿真结果表明, 相对于参比模型, PSO-LSSVM 能够获得更高的网络流量预测精度, 更能准确描述网络流量变化规律。

**关键词:** 预测模型; 网络流量; 粒子群优化算法; 最小二乘支持向量机

## Network Traffic Prediction Method Based on Particle Swarm Algorithm Optimizing Least Square Support Vector Machine

LIU Chun

(Network Management Center, Sichuan College of Architectural Technology, Deyang 618000, China)

**Abstract:** Network traffic had long related and nonlinear characteristics, in order to improve the prediction accuracy of network traffic, this paper proposed a network traffic prediction method based on particle swarm algorithm optimizing the parameters of least square support vector machine. Parameters of least square support vector machine were taken as the position vector of particle, and then the particle swarm algorithm is used to find the optimal parameters of the model, finally, the prediction model of traffic model is established based on least square support vector machine with the optimal parameters. The simulation results showed that the proposed model had improved prediction accuracy ompared with other network traffic prediction models and could more accurately describe the change rule of network traffic.

**Key words:** prediction model; network traffic; particle swarm optimization; least square support vector machine

随着计算机用户增加, 网络交流日益频繁, 网络管理工作越来越繁重。近年来, 进行网络流量预测研究受到了专家学者的广泛关注<sup>[1]</sup>。

网络流量数据在真实的环境中呈现出相当明显的非线性与多尺度特性, 使得传统的网络流量模型如 Poisson 模型、AR 模型、ARMA 模型以及 Markov 模型等已不能有效地对网络流量进行逼近、预测<sup>[2,3]</sup>。近年来一些能够刻画网络流量长相关、非线性等特性的预测模型得到了较快的发展, 如灰色模型、神经网络模型(ANN)、支持向量机模型(SVM)等<sup>[4]</sup>。灰色模型的非线性处理能力不及 ANN 模型与 SVM 模型, ANN 网络结构难以确定<sup>[5]</sup>。基于统计理论的 SVM 模型能够较好地解决小样本、非线性等问题, 在网络流量预测中

优于 ANN 模型<sup>[6]</sup>。最小二乘支持向量机(LSSVM)在 SVM 的基础上有效地避免了求解耗时的二次规划问题, 加快了学习速度, 并且具有较好的泛化能力。LSSVM 预测的性能与参数的选择密切相关, 为了获得更好的网络流量预测结果, 应对 LSSVM 参数进行有效的优化, 不同的专家学者采用不同的方法进行 LSSVM 参数优化, 以提高网络流量的预测精度<sup>[7,8]</sup>。粒子群优化算法(particle swarm optimization, PSO)是一种智能群体搜索算法, 具有很强的全局搜索能力, 而且非常适合于对 LSSVM 的参数进行优化<sup>[9]</sup>。

考虑到网络流量具有流量长相关、非线性等特性, 为了提高网络流量的预测精度, 提出一种粒子群算法优化最小二乘支持向量机参数的网络流量预测模型

<sup>①</sup> 收稿时间:2014-02-19;收到修改稿时间:2014-04-01

(PSO-LSSVM). 首先将最小二乘支持向量机参数作为粒子的位置向量, 然后利用粒子群算法找到模型的最优参数, 最后采用最优参数最小二乘支持向量机建立网络流量预测模型. 仿真结果表明, 相对于参比预测模型, PSO-LSSVM 模型能够获得更高的网络流量预测精度, 更能准确描述网络流量变化规律, 取得了比较满意的结果.

### 1 最小二乘支持向量机

对一组样本集  $\{(x_i, y_i)\}_{i=1}^m$ ,  $x_i \in R^n$  表示输入向量,  $y_i \in R$  表示相应的输出,  $m$  表示样本规模, 通过一个非线性映射函数  $\varphi(\cdot)$  将样本映射到高维特征空间, 然后进行线性回归, 可用式(1)表示:

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

式中,  $w$  表示特征空间的权值向量,  $b$  表示偏置量<sup>[10]</sup>.

根据结构风险最小化原则, 利用LSSVM进行函数回归时优化, 目标见式(2):

$$\begin{aligned} \min J(w, \xi) &= \frac{1}{2} \|w\|^2 + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2 \\ \text{s.t.} & \\ y_i &= w \cdot \varphi(x) + b + \xi_i, (i=1, 2, L, m) \end{aligned} \quad (2)$$

式中,  $\gamma$  表示误差惩罚参数;  $\xi_i$  表示松弛变量.

通过引入拉格朗日乘子(Lagrange multiplier)将式(3)的约束优化问题转变为无约束对偶空间优化问题, 见式(4):

$$\begin{aligned} L(w, b, \xi, \alpha) &= \frac{1}{2} \min \|w\|^2 + \gamma \sum_{i=1}^n \xi_i^2 \\ &\quad - \sum_{i=1}^m \alpha_i (w^T \varphi(x_i) + b + \xi_i - y_i) \end{aligned} \quad (3)$$

式中,  $\alpha_i$  表示拉格朗日乘子.

根据KKT条件:

$$\begin{cases} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^m \alpha_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^m \alpha_i \varphi(x_i) \\ \frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = C \xi_i \\ \frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \varphi(x_i) + b + \xi_i - y_i = 0 \end{cases} \quad (4)$$

消去特征空间的权值向量  $w$  和松弛变量  $\xi_i$ , 得到

式(5):

$$\begin{bmatrix} 0 & Q^T \\ Q & K + C^{-1}I \end{bmatrix} \begin{bmatrix} b \\ A \end{bmatrix} = \begin{bmatrix} 0 \\ Y \end{bmatrix} \quad (5)$$

根据Mercer条件, 核函数定义见式(6):

$$K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j) \quad (6)$$

基于径向基核函数性能一般要优于其它核函数, 径向基核函数定义见式(7):

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (7)$$

式中,  $\sigma$  表示径向基核函数宽度.

最后得到LSSVM网络预测模型:

$$f(x) = \sum_{i=1}^m \alpha_i \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) + b \quad (8)$$

从式(8)可知: LSSVM的学习和泛化能力受到 $\gamma$ ,  $\sigma$  影响. 为了获得更优的核函数参数 $\sigma$ 和正则化参数 $\gamma$ , 减少主观经验选择参数的盲目性, 采用粒子群算法对参数 $\gamma$ ,  $\sigma$ 进行优化.

## 2 PSO-LSSVM的网络流量预测模型

### 2.1 粒子群算法

PSO 算法是一种用来模拟鸟群寻找栖息地行为的群体智能优化算法, 每个粒子代表一个可能解向量, 粒子优劣由适应度函数决定, 并通过向个体和全局最优解学习来不断更新自身位置与速度, 从而达到全局寻优的目的. 粒子的位置和速度分别为  $X_i=(x_{i1}, x_{i2}, \dots, x_{id})$  和  $V_i=(v_{i1}, v_{i2}, \dots, v_{id})$ , 它们的更新策略为:

$$V_{id}^{(t+1)} = \omega \cdot V_{id}^{(t)} + c_1 r_1 (p_{id}^{(t)} - X_{id}^{(t)}) + c_2 r_2 (p_{gd}^{(t)} - X_{id}^{(t)}) \quad (9)$$

$$X_{id}^{(t+1)} = X_{id}^{(t)} + V_{id}^{(t+1)} \quad (10)$$

$$\omega = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{t_{\max}} \times t \quad (11)$$

式中,  $p_{id}^{(t)}$  与  $p_{gd}^{(t)}$  分别表示迭代次数 $t$ 时刻粒子的个体和全局最优解;  $\omega$ 表示惯性权重,  $\omega_{\min}$ 与 $\omega_{\max}$ 分别表示 $\omega$ 的最小值与最大值;  $r_1, r_2$ 表示(0, 1)之间的随机数;  $c_1$ 和 $c_2$ 表示学习因子;  $t_{\max}$ 与 $t$ 分别表示最大和当前迭代次数<sup>[11]</sup>.

### 2.2 PSO-LSSVM 网络流量预测模型

LSSVM 的待优化参数为  $\gamma$  和  $\sigma$ , 因此参数优化的

数学模型为:

$$M = \{\gamma, \sigma\} \tag{12}$$

LSSVM 参数优化目的是提高网络流量的预测精度, 因此采用预测精度( $G$ )作为 LSSVM 参数优化目标函数:

$$\begin{aligned} & \max_M G(M) \\ & s.t. \\ & \begin{cases} \gamma_{\min} \leq \gamma \leq \gamma_{\max} \\ \sigma_{\min} \leq \sigma \leq \sigma_{\max} \end{cases} \end{aligned} \tag{13}$$

PSO-LSSVM的网络流量预测工作步骤为:

1) 收集网络流量序列  $\{x_1, x_2, \dots, x_n\}$ , 然后确定最佳延迟时间( $\tau$ )和嵌入维 $m$ 对原始时间序列进行相空间重构, 得到 LSSVM 的学习样本:

$$\mathbf{X} = \begin{bmatrix} x_1 & x_{1+\tau} & \dots & x_{1+(m-1)\tau} \\ x_2 & x_{2+\tau} & \dots & x_{2+(m-1)\tau} \\ \dots & \dots & \dots & \dots \\ x_{n-(m-1)\tau} & x_{n-(m-1)\tau+\tau} & \dots & x_{n-1} \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} x_{1+(m-1)\tau+1} \\ x_{2+(m-1)\tau+1} \\ \dots \\ x_n \end{bmatrix}$$

2) 设定PSO算法所需种群大小为 $M$ , 学习因子 $c_1$ 和 $c_2$ , 惯性权重 $\omega_{\min}$ 与 $\omega_{\max}$ 和最大迭代次数 $t_{\max}$ , 并初始化粒子的位置向量和速度。

3) 将经过相空间重构后的训练本集输入LSSVM进行学习, 根据式(14)计算出每个粒子的适应度值, 并找出个体和全局最优位置 $p_d$ 和 $p_g$ 。

$$fit = \sum_{i=1}^l (y_i - \hat{y}_i)^2 \tag{14}$$

式中,  $y_i$  和  $\hat{y}_i$  分别表示预测值和实际值。

- 4) 更新粒子的速度、位置和惯性权重。
- 5) 计算每个粒子适应度, 并更新 $p_d$ 和 $p_g$ 。
- 6) 若达到最大迭代次数, 保存全局最优粒子位置, 否则, 则将 $t=t+1$ , 跳转至步骤4)。
- 7) 利用最优位置对应的LSSVM参数( $\gamma, \sigma$ )构建网络流量预测模型, 并对测试样本进行预测。

### 3 仿真实验

#### 3.1 数据来源

为了验证 PSO-LSSVM 模型的有效性, 采用实例进行验证, 数据来源于文库: <http://newsfeed.ntcu.net/~news/2013/>的每小时网络访问流量, 共得到 720 个数据, 数据见图 1。

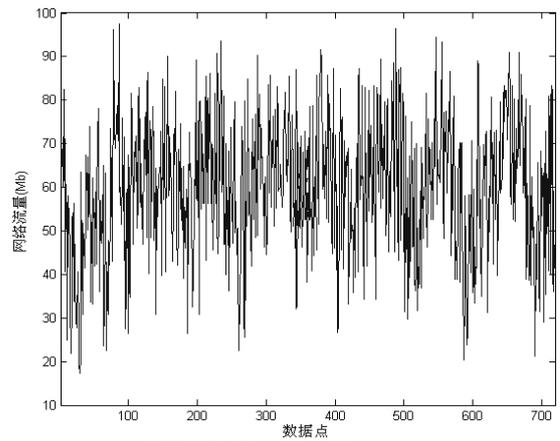


图1 网络流量原始数据

#### 3.2 数据预处理

由于网络流量受到多种因素影响, 数据变化范围较大, 因此, 对网络流量数据进行无量纲化处理, 具体为:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{15}$$

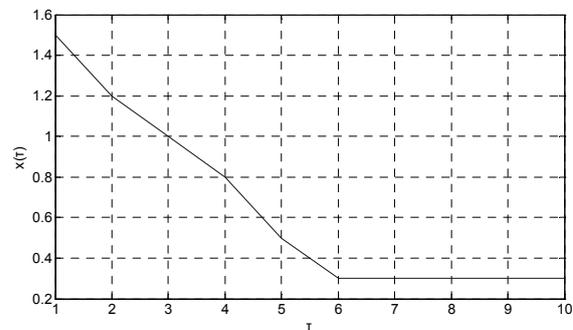
根据式(16)对最终预测结果进行反归一化, 得到网络流量最终预测值:

$$x = x' \times (x_{\max} - x_{\min}) + x_{\min} \tag{16}$$

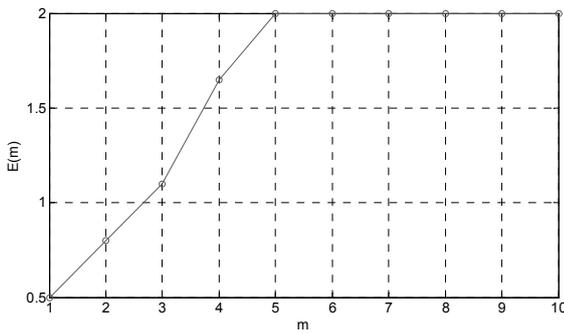
式中,  $x$  为原始值,  $x_{\max}$ 、 $x_{\min}$  分别为最大与最小值。

#### 3.3 学习样本构建

由于网络流量是一维数据, 同时具有混沌性<sup>[12]</sup>, 因此需要采用相空间重构方法对网络流量数据进行重构, 本文采用互信息法对延迟时间( $\tau$ )进行求解, 结果如图2(a)所示。从图2(a)可知, 当 $\tau=6$ , 互信息函数的值达到第一极小值, 所以网络流量时间序列的最优 $\tau=6$ 。采用相关联维法确定嵌入维数为 $m$ , 网络流量时间序列在不同嵌入维数下的关联维数, 如图2(b)所示。从图2(b)可知, 当嵌入维数 $m=7$ 时, 关联维数达到饱和



(a) 计算延迟时间( $\tau$ )



(b) 计算嵌入维(M)

图 2 计算延迟时间(τ)和嵌入维(M)

状态, 因此最佳嵌入维数为  $m=7$ , 采用  $\tau=6, m=7$  对图 1 的网络流量数据进行重构, 得到 LSSVM 的学习样本.

### 3.4 对比模型与评价指标

为了测试 PSO-LSSVM 模型进行网络流量预测的优越性, 选择标准最小二乘支持向量机(LSSVM)、BP 神经网络(BPNN)作为对比模型. 模型性能评价指标采用相对误差(RE%)、均方根误差(RMSE), 具体定义见下式:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (17)$$

$$RE = \frac{y_i - \hat{y}_i}{y_i} \times 100\% \quad (18)$$

式中,  $y_i$  表示网络流量实际观测值,  $\hat{y}_i$  表示预测值.

### 3.5 模型的实现

将重构的网络流量训练集对LSSVM进行训练, 后 20组数据进行模型检测, 在MATLAB 2012软件平台上, 采用径向基核作为LSSVM模型的核函数, 通过 PSO优选参数得到 $\gamma=3.78, \sigma^2=0.01$ . PSO对LSSVM参数寻优曲线见图3.

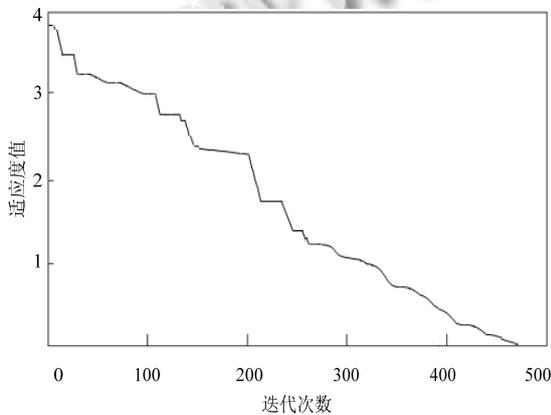


图3 PSO对LSSVM参数寻优曲线

### 3.6 结果与分析

PSO-LSSVM、BPNN 与 LSSVM 模型对检测样本的预测结果见图 4. 从图 4 可知, 各模型的预测结果均比较接近观测值, 主要是因为这 3 个模型均具有较强的非线性逼近能力, 能够很好地对具有较高非线性的网络流量数据进行逼近, 但是 PSO-LSSVM 模型的预测结果与观测值更为贴近, 主要是由于利用 PSO 算法能够找到更优的 LSSVM 模型参数.

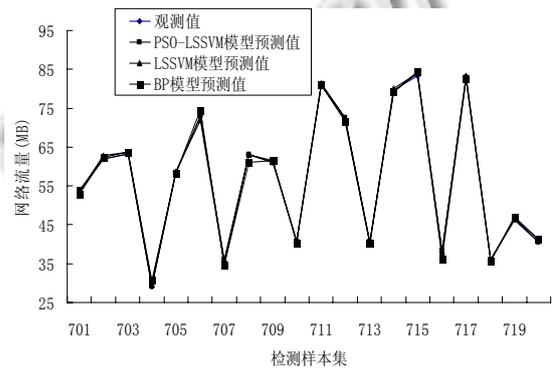


图4 不同模型的网络流量预测结果

各模型的性能评价指标见表 1. 从表 1 可知, PSO-LSSVM 预测模型的 20 个检测样本中, 所有时段点的相对误差的绝对值均小于 2%, 这充分说明 PSO-LSSVM 预测模型具有较高的预测精度. 对于 LSSVM 预测模型, 有 2 个时段点的相对误差的绝对值在 3%以上. 而在 BPNN 预测模型的预测结果中, 有 4 个时段点的相对误差绝对值较大, 分别是 704、707、712、716 号样本, 其预测的相对误差分别为 8.80%、6.24%、4.04%、6.52%, 这充分说明了 BPNN 的预测精度低于 PSO-LSSVM 和 LSSVM.

从表 1 还可看到, PSO-LSSVM 预测模型的 RMSE 为 0.52, 而 LSSVM、BPNN 的 RMSE 分别为 1.10 和 1.58. 从各模型的性能评价指标可知, 相对于参比模型, PSO-LSSVM 网络流量预测模型具有更高的预测精度.

表 1 预测集的仿真预测结果及相对误差

样本编号	实际值	PSO-LSSVM		LSSVM		BPNN	
		预测值	RE%	预测值	RE%	预测值	RE%
701	53.65	53.97	-0.60	53.94	-0.55	52.13	2.82
702	61.25	61.88	-1.03	60.17	1.76	60.57	1.11
703	63.58	63.24	0.53	63.95	-0.59	62.63	1.49

704	28.89	29.04	-0.52	28.14	2.58	26.35	8.80
705	58.40	58.19	0.36	56.60	3.08	59.44	-1.78
706	73.04	72.92	0.16	74.48	-1.97	71.62	1.94
707	35.73	35.98	-0.71	35.36	1.02	33.50	6.24
708	62.99	63.00	-0.01	62.51	0.77	63.65	-1.04
709	60.78	61.57	-1.30	59.14	2.70	61.20	-0.69
710	41.04	40.75	0.71	41.61	-1.40	40.17	2.13
711	81.46	80.93	0.65	82.31	-1.04	82.71	-1.53
712	71.41	72.61	-1.69	73.04	-2.29	68.52	4.04
713	40.02	40.65	-1.58	39.49	1.33	41.15	-2.82
714	79.62	79.17	0.56	80.71	-1.37	79.84	-0.28
715	83.72	83.93	-0.25	82.20	1.82	84.34	-0.74
716	38.04	38.33	-0.76	36.82	3.21	35.56	6.52
717	82.10	82.93	-1.00	82.56	-0.55	80.94	1.42
718	36.79	36.10	1.87	37.79	-2.73	38.08	-3.53
719	46.48	46.12	0.78	47.62	-2.45	45.67	1.75
720	40.99	40.63	0.87	39.44	3.78	38.02	7.25
RMSE		0.52		1.10		1.58	

#### 4 结语

在利用 LSSVM 进行网络流量预测时,其预测性能受到参数的影响,若参数选取不恰当,则预测精度难以达到相关要求,因此,模型的参数选择至关重要,为了获得更优的 LSSVM 参数,在研究中,利用 PSO 算法对 LSSVM 参数进行优化,得到了 LSSVM 的最优参数,仿真结果表明,相对于参比模型,PSO-LSSVM 模型的网络流量预测精度更高,是一种有效的网络流量预测模型,在实际应用中具有较高的参考价值。

#### 参考文献

- 1 Nguyen TT, Armitage G. A survey of techniques for internet traffic classification using machine learning. *IEEE Communications Surveys and Tutorials*, 2008, 10(4): 56–76.
- 2 Silva CG. Time series forecasting with a nonlinear model and the scatter search meta-heuristic. *Information Sciences*, 2008, 178(16): 3288–3299.
- 3 姚奇富,李翠凤,马华林,张森.灰色系统理论和马尔柯夫链相结合的网络流量预测方法. *浙江大学学报(理学版)*, 2007, 34(4): 396–400.
- 4 Este A, Gringoli F, Salgarelli L. Support vector machines for TCP traffic classification. *Computer Networks*, 2009, 53(14): 2476–2490.
- 5 李超,赵海,葛新,张君.基于混沌特征的网络延迟预测模型. *电子学报*, 2009, 37(12): 2657–2661.
- 6 张强,许少华,李盼池.分式过程神经网络在网络流量预测中的应用. *计算机工程与应用*, 2012, 48(35).
- 7 Huang J. Electromechanical equipment state forecasting based on genetic algorithm–support vector regression. *Expert Systems with Applications*, 2011, 38(7): 8399–8402.
- 8 罗赘骞,夏靖波,王焕彬.混沌-支持向量机回归在流量预测中的应用研究. *计算机科学*, 2009, 36(7): 244–246.
- 9 赵云,肖鬼,陈阿林.基于加权支持向量回归的网络流量预测. *计算机工程与应用*, 2012, 48(21): 103–106.
- 10 潘岚,王仲君.基于特征指数加权的最小二乘支持向量机算法. *计算机系统应用*, 2012, 21(5): 205–208.
- 11 李欣然,靳雁霞.权重自适应调整的混沌量子粒子群优化算法. *计算机系统应用*, 2012, 21(8): 127–130.
- 12 温祥西,孟相如,马志强,张永春.小时间尺度网络流量混沌性分析及趋势预测. *电子学报*, 2012, 40(8): 1609–1616.