

GAFSA 优化 SVR 的网络流量预测模型研究^{*}

王瑞雪, 刘 渊

(江南大学 数字媒体学院, 江苏 无锡 214122)

摘要: 现有的诸多网络流量预测模型存在预测稳定性不好、精度较低等问题。针对此类问题, 研究了一种通过 GAFSA(全局人工鱼群算法)优化 SVR 模型的网络流量预测方法。GAFSA 是一种群智能优化算法, 寻优效果显著。采用 GAFSA 对 SVR 预测模型进行参数寻优, 可以得到使预测效果最佳的训练参数; 使用这组最优参数训练 SVR, 建立网络流量预测模型, 可以很好地改善基于其他智能优化算法改进的 SVR 网络流量预测模型多次预测结果相差较大的问题, 使预测结果趋于稳定, 同时也可以提高预测精准度。仿真结果表明, GAFSA-SVR 网络流量预测模型与其他模型相比, 预测结果基本稳定, 精准度提高到 89% 以上, 对于指导网络控制行为、分析网络安全态势有重要意义。

关键词: 网络流量预测; 参数优化; 支持向量回归机; 全局人工鱼群算法; 自相似性

中图分类号: TP393 **文献标志码:** A **文章编号:** 1001-3695(2013)03-0856-05

doi: 10.3969/j.issn.1001-3695.2013.03.055

Study on network traffic forecast model of SVR optimized by GAFSA

WANG Rui-xue, LIU Yuan

(School of Digital Media, Jiangnan University, Wuxi Jiangsu 214122, China)

Abstract: There are some problems, such as low precision, on existed network traffic forecast model. In accordance with these problems, this paper proposed the network traffic forecast model of support vector regression (SVR) algorithm optimized by global artificial fish swarm algorithm (GAFSA). GAFSA constituted an improvement of artificial fish swarm algorithm, which was a swarm intelligence optimization algorithm with a significantly effect of optimization. The optimum training parameters could be calculated with optimizing by chosen parameters, which would make the forecast more accurate. With the optimum training parameters searched by GAFSA algorithm, a model of network traffic forecast, which greatly solved problems of great errors in SVR improved by others intelligent algorithms, could be built with the forecast result approaching stability and the increased forecast precision. The simulation shows that, compared with other models, the forecast results of GAFSA-SVR network traffic forecast model is more stable with the precision improves to more than 89%, which plays an important role on instructing network control behavior and analyzing security situation.

Key words: network traffic forecast; optimization of parameters; support vector regression (SVR); global artificial fish swarm algorithm (GAFSA); self-similarity

0 引言

一个有精确预测能力的网络流量预测模型, 对于理解即将要发生的网络行为、分析网络安全状况、指导网络安全检测与控制具有重要的意义。目前, 随着网络宽带化步伐加快, 流量数据逐渐变得巨大, 数据变化也逐渐复杂, 管理难度增高。找寻更准确、稳定的网络流量预测模型已经成为众多研究热点之一。

现有的网络流量预测模型可分为线性、非线性两种。典型的线性模型包括基于短相关特性的自回归 (auto regressive, AR) 模型^[1]、自回归滑动平均 (auto regressive moving average, ARMA) 模型^[2,3]、差分自回归滑动平均 (auto regressive integrated moving average, ARIMA) 模型^[4,5]和基于长相关特性的差分自回归求和滑动平均模型 (fractional auto regressive integrated moving average, FARIMA)^[6,7]等。单一的线性模型虽然简单易

实现, 但逐渐难以描述和预测日益复杂多变的网络流量特性。典型的非线性模型包括灰色模型^[8,9]、小波预测模型^[10]、神经网络预测模型^[11~14]、支持向量机回归预测模型 (support vector regression, SVR)^[15], 以及针对这些模型融合和改进的各类模型。SVR 预测模型泛化能力强^[16], 能够寻找到全局最优解, 并且避免了“维数灾难”, 预测效果优于其他非线性模型, 得到了广泛的应用^[17~19]。当前已有一些针对 SVR 预测模型的改进研究。主要包括对 SVR 预测模型自身的改进和对 SVR 预测模型参数选择过程的改进。在 SVR 模型的预测过程中, 参数的选择对该模型的预测效果起着关键的作用。目前常用于 SVR 参数选择的智能算法主要是遗传算法 (genetic algorithm, GA) 和粒子群优化算法 (particle swarm optimization, PSO)。GA 优化的 SVR 预测模型 (GA-SVR)^[20,21] 参数设置复杂, 该算法的寻优方式由概率决定, 结果很不稳定。PSO 优化的 SVR 模型 (PSO-SVR)^[22] 容易陷入局部极值。根据 PSO-SVR 改进的混

收稿日期: 2012-08-06; **修回日期:** 2012-09-16 **基金项目:** 江苏省自然科学基金重点研究专项基金资助项目 (BK2011003); 国家自然科学基金资助项目 (61103223); 江苏省六大人才高峰基金资助项目

作者简介: 王瑞雪 (1987-), 女, 硕士, 主要研究方向为计算机网络流量预测 (121654971@qq.com); 刘渊 (1967-), 男, 教授, 硕士, 主要研究方向为计算机网络安全、数字媒体。

沌粒子群算法 (chaos particle swarm optimization, CPSO) 优化 SVR 预测模型 (CPSO-SVR)^[23] 提高了 PSO-SVR 模型的预测精度,但预测的稳定性依然不高。

人工鱼群算法^[24] (artificial fish swarm algorithm, AFSA) 由李晓磊博士于 2002 年提出,是一种现代启发式搜索算法,寻优效果佳,具有并行性、简单性、很快能跳出局部极值等特点。该模型鲁棒性强、全局收敛性好,对于初值敏感性小。江铭炎等人^[25] 提出一种 AFSA 算法的改进算法——全局人工鱼群算法 (global artificial fish swarm algorithm, GAFSA), 将全局信息加入寻优过程,使鱼群算法的寻优效果更好,收敛性更强。本文研究结合全局人工鱼群算法的新颖的寻优策略对 SVR 预测模型进行改进,使用 GAFSA 算法对 SVR 算法进行参数寻优,并用寻找到的最优参数训练 SVR,建立 GAFSA-SVR 预测模型,预测网络流量。

1 基本 SVR 回归预测模型

SVR 算法的基本思想是:利用一个非线性映射,把数据映射到一个高维的特征空间里,免去了在高维空间计算复杂的点积运算。其基本过程如下:

数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l)\}$, x_i 为输入值, y_i 为对应的预测值, $i = 1, \dots, l$ 。设定回归函数为 $f(x) = w \times \varphi(x) + b$ 。其中: w 是自回归系数,即权重向量; b 是误差值。通过风险最小化正则函数训练参数 w 和 b , 确定函数 $f(x)$ 为

$$\min \frac{1}{2} \|w\|^2 + C \frac{1}{k} \sum_{i=1}^k \varepsilon(f(x_i) - y_i) \quad (1)$$

$$\text{s. t. } \varepsilon(f(x_i) - y_i) = \begin{cases} |f(x_i) - y_i| - \varepsilon & |w \times \varphi(x) + b - y_i| \geq \varepsilon \\ 0 & |w \times \varphi(x) + b - y_i| < \varepsilon \end{cases}$$

其中: $\varepsilon(\cdot)$ 为 Vapnik 等人提出的 ε 不敏感损失函数, ε 为 ε -intensive 损失参数; C 为惩罚因子,用来平衡正经验风险和置信范围,使风险最小化。

将寻找最优超平面的二次规划问题转换为

$$\min_{w, b, \xi_i, \xi_i^*} \frac{1}{2} \|w\|^2 + c \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (2)$$

$$\text{s. t. } \begin{cases} y_i - w \times \varphi(x) - b \leq \varepsilon + \xi_i & \xi_i \geq 0, i = 1, 2, \dots, l \\ w \times \varphi(x) + b - y_i \leq \varepsilon + \xi_i^* & \xi_i^* \geq 0, i = 1, 2, \dots, l \end{cases}$$

其中: $w \in \mathbb{R}^n$, $b \in \mathbb{R}$; φ 为将输入数据从低维的非线性回归问题映射到高维特征空间转换为线性问题所做的映射; ε 和 ε^* 为非负的松弛变量。

引入朗格拉日乘子 a_i 和 a_i^* , 将式(3)转换为对偶问题:

$$\min_{a_i^{(*)} \in \mathbb{R}^{2l}} \frac{1}{2} \sum_{i,j=1}^l (a_i^* - a_i) (a_j^* - a_j) K(x_i, x_j) +$$

$$\varepsilon \sum_{i=1}^l (a_i^* + a_i) - \sum_{i=1}^l y_i (a_i^* - a_i) \quad (3)$$

$$\text{s. t. } \begin{cases} \sum_{i=1}^l (a_i^* - a_i) = 0 \\ 0 \leq a_i^{(*)} \leq C, i = 1, 2, \dots, l \\ 0 \leq a_i^{(*)} \leq C \quad i = 1, 2, \dots, l \end{cases}$$

其中: $K(x_i, x_j)$ 是核函数。最终得到回归函数 $f(x)$ 的形式为

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) (\varphi(x_i), \varphi(x)) + b \quad (4)$$

2 GAFSA 算法优化的 SVR 预测模型

2.1 SVR 预测模型相关参数

不敏感系数 ε 、惩罚系数 C 以及核函数的选择是 SVR 预

测模型参数选择的重要内容。常用的核函数有线性核、多项式核、径向基 (radial basis function, RBF) 核、Sigmoid (S 形) 核、傅里叶级数核、样条核等。其中, RBF 核是应用最广泛的核函数,通过参数的选择,可以适用于任何样本。本文选择 RBF 核作为 SVR 预测模型的核函数。RBF 核函数的形式为

$$K(x, x_i) = \exp \left[-\frac{\|x - x_i\|^2}{\sigma^2} \right] \quad (5)$$

RBF 核的参数 σ 隐含地决定了数据映射到新空间的分布。

综上所述, SVR 模型的预测效果由 RBF 参数 σ 、不敏感系数 ε 和惩罚系数 C 决定。

2.2 基本人工鱼群算法

人工鱼群算法是一种基于动物行为的群体智能优化算法,该算法通过模拟鱼类的觅食、聚群、追尾、随机等行为在搜索域中进行寻优。

觅食行为是指设一条人工鱼 i 当前的状态为 X_i , visual 为其视野范围, step 是移动步长。在该鱼视野范围内随意选择一个状态 X_j :

$$X_j = X_i + \text{visual} \times \text{rand}() \quad (6)$$

若该状态的解优于 X_i 状态的解,则向该方向前进一步:

$$X_i^{t+1} = X_i^t + \frac{X_j - X_i^t}{\|X_j - X_i^t\|} \times \text{step} \times \text{rand}() \quad (7)$$

若不满足,则返回重新选择状态 X_j 。尝试 try_number 次后仍不满足,执行随机行为。

群聚行为设定为:人工鱼在游动过程中探索当前邻域 (visual) 内的伙伴数目 n_f 及中心位置 X_c , X_c 的解为 Y_c , δ 为拥挤因子,若 $Y_c n_f < \delta Y_i$ (因最大值问题和最小值问题可以互相转换,以下均以最小值问题讨论),则朝伙伴中心移动。

$$X_i^{t+1} = X_i^t + \frac{X_c - X_i^t}{\|X_c - X_i^t\|} \times \text{step} \times \text{rand}() \quad (8)$$

追尾行为是指人工鱼在游动过程中探索视野范围内解 Y_j 最优的伙伴 X_j , 若 $Y_c n_f < \delta Y_i$, 则朝 X_j 移动:

$$X_i^{t+1} = X_i^t + \frac{X_j - X_i^t}{\|X_j - X_i^t\|} \times \text{step} \times \text{rand}() \quad (9)$$

随机行为是人工鱼随意在视野范围内选择一个状态,向该方向移动:

$$X_i^{t+1} = X_i^t + \text{visual} \times \text{rand}() \quad (10)$$

鱼类通过比较这四种行为的解,选择解最优的方式移动;达到设定的迭代次数或结束条件,算法结束。最终得到最优解。

2.3 全局人工鱼群算法

为了提高人工鱼群算法的全局搜索能力,提出将全局最优信息加入人工鱼的位置更新模式中,即为全局人工鱼群算法。

在觅食行为、群聚行为、追尾行为中,将全局最优人工鱼 $X_{\text{best_af}}$ 信息加入其前进方向中,设在这三种行为中选择的行进方向的状态为 X_j , 则移动的方向改为向 X_j 位置和全局最优位置 $X_{\text{best_af}}$ 的向量和方向前进一步:

$$X_i^{t+1} = X_i^t + \left(\frac{(X_j - X_i^t) + (X_{\text{best_af}} - X_i^t)}{\|(X_j - X_i^t) + (X_{\text{best_af}} - X_i^t)\|} \right) \times \text{step} \times \text{rand}() \quad (11)$$

这样,就克服了基本人工鱼算法位置更新中只有局部信息没有全局信息的位置更新模式,使算法更易达到全局最优解,收敛速度加快,精度提高。

2.4 GAFSA-SVR 模型设计

设计 GAFSA-SVR 模型为一个三元自变量求最小值的过

程,变量即为要选择的参数 C, σ, ε , 所求的最小值即为适应度函数的最小值。将自变量作为参数代入 SVR 模型训练并预测, 采用预测结果的相对误差总和作为适应度评估函数。设 n 为要预测的流量数据的数目, y_i 为真实数据, y_i^* 为预测数据, 则适应度函数为

$$\text{fitness} = \sum_{i=1}^n |y_i - y_i^*| \quad (12)$$

使适应度最小的自变量即为要寻优的参数 C, σ, ε 。采用寻优得到的参数 C, σ, ε 作为 SVR 的参数训练预测模型, 进行网络流量预测。

3 网络流量的自相似性

如果一组数据具有较强的长记忆性, 即过去的数据和现在及未来的数据具有一定的自相似性, 就可以在一定程度上利用过去和现在的数据来预测未来的数据。若网络流量也满足这种特性, 则可通过过去的流量数据来预测未来的网络状况。

1993 年起, Leland 等人^[26] 在传统的泊松过程基础上开始研究将自相似的概念引入到网络流量特性的研究当中; 随后有很多相关的研究^[27] 均表明网络业务流量的突发过程更符合自相似模型。

3.1 自相似性度量

自相似性的强度可以用 Hurst 参数来估计, 用 H 来表示。Hurst 参数产生于应用数学, 广泛应用于分形和混沌理论、长记忆过程、谱分析等数学领域当中。Hurst 指数是这样衡量时间序列的:

a) 当 $0 \leq H < 0.5$ 时, 序列表现出均值回复, 倾向说反持久性、遍历性, 突变性和波动性高于随机序列;

b) 当 $H = 0.5$ 时, 序列近似于高斯独立分布的随机过程, 变量间相互独立。

c) 当 $0.5 < H \leq 1$ 时, 变量之间具有较强的相关性, 序列存在较强的长记忆性, 表现出趋势追随的特性, 即具有自相似性。

3.2 Hurst 参数的估算方法

现有的估算 Hurst 参数的算法有很多, 包括方差—时间图法、绝对值法、残差方差法、重标极差分析法 (rescaled range analysis, R/S)、周期图法、Whittle 法、小波分析法 (AV 法) 等, 均能较好地估算出 Hurst 参数的大小。

R/S^[28,29] 分析法通常用来分析时间序列的分形特征和长期记忆过程, 最初由英国水文学家赫斯特在研究尼罗河水坝工程时提出, 后来它被用在各种时间序列的分析之中。R/S 方法因为其强健性一直受到研究者的青睐, 迄今为止是估计 Hurst 参数最常用的方法。本文采用 R/S 方法估计采集数据的 Hurst 参数, 来研究采集数据的自相关性, 进而验证该组流量数据的可预测性。

4 仿真实验

4.1 流量数据的采集与样本集的建立

4.1.1 流量采集

采集了 <http://news.ntu.edu.tw/stat/> 从 2012 年 3 月 1 日 0 时到 2012 年 3 月 13 日 13 时 Incoming articles 每小时的流量, 共得到 301 个流量数据, 组成 301×1 的原始样本集 data。本文是以小时为时间尺度来进行网络流量预测的。

$$\text{data} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{301} \end{bmatrix} \quad (13)$$

4.1.2 可预测性验证

本文采用 R/S 法估计这组数据的 Hurst 参数。定义时间序列 $X_i = \{X_1, X_2, \dots, X_n\}$ 中:

$$W_j = (X_1 + X_2 + \dots + X_j) - j\bar{X}(n) \quad j=1, 2, 3, \dots, n \quad (14)$$

$$R(n) = \max(0, W_1, W_2, \dots, W_n) - \min(0, W_1, W_2, \dots, W_n) \quad (15)$$

这样就获得了这组数据的 R 值 $R(n)$ 。 $S(n)$ 为这组数据的标准差。Hurst 指出, 很多自然的随机序列, 当 n 趋向于无穷时, R/S 近似服从以下规律:

$$E\left[\frac{R(n)}{S(n)}\right] \approx n^H \quad (16)$$

假定输入时间序列为 X_i , 长度为 N_i , R/S 法的步骤如下:

a) 取定一组 n 值, $n = \{10, 11, \dots, N/2\}$, 设 n 的第 i 值为 n_i 。

b) 拆分 X_i 为 $\lfloor N/n_i \rfloor$ 个互不重叠的数据组, 各包含 n_i 个数据, 分别计算每组 R 值与 S 值, 可得一组 R/S 值, 并求这组 R/S 的均值 RS_i 。

c) 重复步骤 b) $N/2 - 10 + 1$ 次。

d) 使用最小二乘法拟合曲线 $\log RS_i = \log c + k \log n_i$, 求得其斜率 k , 这组数据的 Hurst 参数 H 即为斜率 k 。

通过上述方法, 求得样本时间序列 data 的 Hurst 参数 $H = 0.7012$, 落入区间 $(0.5, 1]$, 说明序列 data 存在长记忆性, 具有较强的自相似性, 即该序列是可预测的。

4.1.3 样本集的建立

对数据 data 进行空间重构。选择嵌入维度为 1, 即使用前一个小时的流量值 X 预测下一个小时的流量 Y 。第 1 ~ 300 个样本作为输入集 X , 第 2 ~ 301 个样本作为输出集 Y , 将原始样本集 data 重组为一组 300×2 的样本集 Data。

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{300} \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_{300} \end{bmatrix} \quad (17)$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{300} \end{bmatrix} = \begin{bmatrix} d_2 \\ d_3 \\ \vdots \\ d_{301} \end{bmatrix} \quad (18)$$

$$\text{Data} = [X \ Y] = \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_{300} & y_{300} \end{bmatrix} = \begin{bmatrix} d_1 & d_2 \\ d_2 & d_3 \\ \vdots & \vdots \\ d_{300} & d_{301} \end{bmatrix} \quad (19)$$

使用前 280 组数据作为训练数据, 后 20 组数据作为测试数据来验证模型的预测效果。

4.2 对比模型及参数设定

分别以 CPSO-SVR 流量预测模型和 GA-SVR 流量预测模型对比验证 GAFSA-SVR 模型的优势性。

GAFSA-SVR 模型设定人工鱼数目为 20, 迭代数为 20, 试探次数 try_number 为 50, 感知距离为 3, 拥挤度因子为 0.5, 步长为 1.5。

CPSO-SVR 模型设定粒子数为 20, 迭代数目为 20, 代表局部搜索能力的参数为 1.5, 代表全局搜索能力的参数为 1.7, 速

率与自变量的比值为 0.5。

GA-SVR 模型也设定种群数量为 20,进化代数数为 20。

4.3 实验效果评价标准

为了评价模型的性能,采用三种评价标准:平均绝对误差 MAE(mean absolute error,单位与所采集的数据同为 MB)、均方根误差 RMSE(root mean square error,单位与所采集的数据同为 MB)、预测准确率 accuracy(单位为%)。MAE 描述了模型整体的误差情况,RMSE 的大小衡量了模型的离散程度,accuracy 衡量了该模型预测的精准程度。MAE 和 RMSE 越低、accuracy 越高,模型的预测效果越好、越稳定。

$$MAE = \frac{1}{n} \sum_{k=1}^n |x_t - \hat{x}_t| \quad (20)$$

$$RMSE = \sqrt{\frac{\sum_{k=1}^n (x_k - \hat{x}_k)^2}{n}} \quad (21)$$

$$accuracy = (1 - \sum_{i=1}^n \frac{|y_i - y_i^*|}{y_i}) \times 100\% \quad (22)$$

4.4 实验结果

这三种模型均含有随机成分,因此,本文分别使用这三种模型对采集的同一组网络流量数据进行预测实验。为验证各模型预测结果的稳定性,每种模型分别进行 10 次实验,并记录每次实验 MAE、RMSE、accuracy 的平均值和 accuracy 最高的一次实验及 accuracy 最低的一次实验。第一组使用 GA-SVR 模型,第二组使用 CPSO-SVR 模型,第三组使用 GAFSA-SVR 模型。三种模型 10 次实验寻优得到的参数及预测结果分别如表 1~3 所示。三种模型 10 次实验的 MAE、RMSE、accuracy 平均值如表 4 所示。

表1 GA-SVR 模型参数寻优及预测结果

实验	C	σ	ε	MAE	RMSE	accuracy
1	26.407	11.497	3.853	2.260	2.745	75.766
2	1.427	3.865	1.889	1.969	2.288	78.921
3	3.355	45.102	7.730	1.893	2.240	79.569
4	19.030	14.101	3.462	2.325	2.778	75.025
5	1.846	13.527	1.188	1.961	2.284	78.963
6	30.116	1.674	4.099	2.132	2.595	77.897
7	4.057	14.541	3.011	2.358	2.755	74.792
8	20.555	3.992	5.229	2.296	2.866	75.632
9	0.418	4.348	9.104	1.341	1.705	85.296
10	5.543	43.656	4.964	2.208	2.566	76.217
最优	0.418	4.348	9.104	1.341	1.705	85.296
最差	4.057	14.541	3.011	2.358	2.755	74.792

表2 CPSO-SVR 模型参数寻优及预测结果

实验	C	σ	ε	MAE	RMSE	accuracy
1	0.232	0.465	0.380	1.508	1.912	83.381
2	0.102	0.741	2.474	1.212	1.346	87.629
3	0.274	1.524	1.731	1.332	1.695	85.255
4	0.288	0.048	0.444	1.570	1.978	82.703
5	0.374	0.925	1.684	1.722	2.137	81.119
6	1.711	1.773	1.721	1.926	2.308	79.255
7	0.163	0.018	1.173	1.171	1.467	87.447
8	0.245	1.078	0.916	1.318	1.675	85.404
9	0.179	0.310	0.270	1.364	1.725	84.935
10	0.326	0.752	1.071	1.609	2.028	82.294
最优	0.102	0.741	2.474	1.212	1.346	87.629
最差	1.711	1.773	1.721	1.926	2.308	79.255

表3 GAFSA-SVR 模型参数寻优及预测结果

实验	C	σ	ε	MAE	RMSE	accuracy
1	0.278	3.073	0.963	0.973	1.226	89.329
2	0.256	3.750	3.625	0.938	1.144	89.888
3	0.254	3.613	1.074	0.937	1.150	89.877
4	0.301	4.726	2.944	0.962	1.161	89.590
5	0.269	4.135	2.134	0.941	1.140	89.856
6	0.266	4.153	1.111	0.939	1.133	89.898
7	0.251	3.545	1.095	0.936	1.151	89.882
8	0.250	3.494	1.205	0.936	1.152	89.883
9	0.250	3.509	2.794	0.936	1.152	89.886
10	0.253	3.556	3.128	0.939	1.153	89.852
最优	0.266	4.153	1.111	0.939	1.133	89.898
最差	0.278	3.073	0.963	0.973	1.226	89.329

表4 三种模型的评价标准均值比较

预测模型	MAE 均值	RMSE 均值	accuracy 均值
GA-SVR	2.074	2.482	77.808
CPSO-SVR	1.473	1.827	83.942
GAFSA-SVR	0.944	1.156	89.794

分别选择了这三种模型的 10 次实验中 accuracy 最接近于平均值的一次实验,展示该次实验的结果图,用来对比三种模型的预测效果。

GA-SVR 模型 10 次的实验结果中 accuracy 最接近于平均值的一次实验为实验 6,MAE 为 2.132,RMSE 为 2.595,准确率为 77.897%,该次实验的预测效果如图 1 所示。

CPSO-SVR 模型 10 次的实验结果中 accuracy 最接近于平均值的一次实验为实验 1,MAE 为 1.508,RMSE 为 1.912,准确率为 83.381%,该次实验的预测效果如图 2 所示。

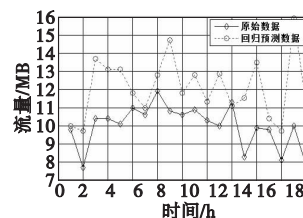


图1 GA-SVR预测效果图(实验6)

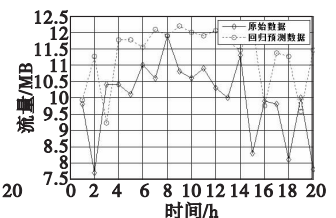


图2 CPSO-SVR预测效果图(实验1)

GAFSA-SVR 模型 10 次的实验结果中 accuracy 最接近于平均值的一次实验为实验 10,MAE 为 0.939,RMSE 为 1.153,准确率为 89.852%,该次实验的预测效果如图 3 所示。

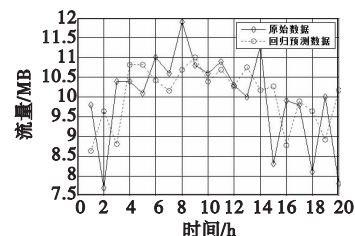


图3 GAFSA-SVR预测效果图(实验10)

4.5 实验结果分析

表 1~3 的对比表明,10 次实验中,CPSO-SVR 模型和 GA-SVR 模型预测效果均不稳定,每次实验预测结果相差较大;而 GAFSA-SVR 模型的预测效果则要稳定得多,10 次的实验结果几乎没有差别,可靠性相当高。

表 4 表明,GAFSA-SVR 网络流量预测模型与 CPSO-SVR 和 GA-SVR 模型相比,10 次实验的平均相对误差较低,均方根误差也较低,预测的准确率较高,均达到 89% 以上。该结果表明,GAFSA-SVR 网络流量预测模型与其他两个模型相比,误差小,稳定度强,且预测准确率高。

图1~3显示了三种模型预测准确率最接近平均值的一次实验的实验效果。通过这三幅图也可以对比出 GAFSA-SVR 模型的优势性。

如上所述,实验结果充分表明,GAFSA-SVR 网络流量预测模型预测精准度高,稳定度好,可靠性强,与 GA-SVR 及 CPSO-SVR 网络流量预测模型相比,具有很强的优势性。

5 结束语

本文研究使用 GAFSA 算法优化 SVR 流量预测模型参数选择过程,建立 GAFSA-SVR 网络流量预测模型。并通过仿真实验,表明了该模型不仅预测准确率高,而且稳定性好,可靠性强,预测效果非常好。未来可以将智能优化参数的算法与 SVR 算法自身的改进过程进行结合,建立更精准可靠的网络流量预测模型。

参考文献:

- [1] 党小超,阎林.基于多元线性自回归模型的流量预测[J].计算机工程,2012,38(1):84-89.
- [2] 邹柏贤,刘强.基于 ARMA 模型的网络流量预测[J].计算机研究与发展,2002,39(12):1645-1652.
- [3] 段智彬,孙恩昌,张延华,等.基于 ARMA 模型的网络流量预测[J].中国电子科学研究院学报,2009,4(4):325-356.
- [4] 薛可,李增智,刘浏,等.基于 ARIMA 模型的网络流量预测[J].微电子学与计算机,2004,21(7):84-87.
- [5] 张冉,赵成龙. ARIMA 模型在网络流量预测中的应用研究[J].计算机仿真,2011,28(1):171-174.
- [6] 李士宁,闫焱,覃征.基于 FARIMA 模型的网络流量预测[J].计算机工程与应用,2006,42(29):148-150.
- [7] 胡玉清,谭献海,宋正阳.基于 FARIMA 的网络建模与性能分析[J].计算机工程与设计,2008,29(18):4666-4714.
- [8] 曹建华,刘渊,戴悦.一种基于灰色神经网络的网络流量预测模型[J].计算机工程与应用,2008,44(5):155-157.
- [9] 马华林,李翠凤,张立燕.基于灰色模型和自适应过滤的网络流量预测[J].计算机工程,2009,35(1):130-152.
- [10] 洪飞,吴志美.基于小波的多尺度网络流量预测模型[J].计算机学报,2006,29(1):166-171.
- [11] 刘杰,黄亚楼.基于 BP 神经网络的非线性网络流量预测[J].计算机应用,2007,27(7):1770-1772.
- [12] 王俊松,高志伟.基于 RBF 神经网络的网络流量建模与预测[J].计算机工程与应用,2008,44(13):7-11.
- [13] 林雪纲,郑成兴,实旻,等.基于 FIR 神经网络的以太网网络流量预测[J].计算机工程,2006,32(8):124-130.
- [14] 侯家利. Elman 神经网络的网络流量预测[J].计算机仿真,2011,28(7):154-157.
- [15] 李丽.基于支持向量机的网络流量预测研究[D].南京:南京理工大学,2007.
- [16] 王定成.支持向量机建模预测与控制[M].北京:气象出版社,2009.
- [17] 谢国强.基于支持向量回归机的股票价格预测[J].计算机仿真,2012,29(4):379-382.
- [18] 陈强,吴慕春,薛月菊,等.支持向量回归机的碳通量预测[J].计算机工程与应用,2009,45(21):235-238.
- [19] 刘显德,高泓.基于 SVR 的组合预测模型及其应用[J].计算机工程与设计,2009,30(19):4506-4523.
- [20] HUANG Ji. Electromechanical equipment state forecasting based on genetic algorithm-support vector regression[J]. Expert Systems with Applications,2011,38(7):8399-8402.
- [21] CHEN Kuan-yu. Forecasting systems reliability based on support vector regression with genetic algorithms[J]. Reliability Engineering and System Safety,2007,92(4):423-432.
- [22] LIN Shih-wei, YING Kuo-ching, CHEN Shih, et al. Particle swarm optimization for parameter determination and feature selection of support vector machines[J]. Expert Systems with Applications,2008,35(4):1817-1824.
- [23] 王治.基于混沌粒子群优化 SVR 的网络流量预测[J].计算机仿真,2011,28(5):151-154.
- [24] 李晓磊,钱积新.人工鱼群算法:自下而上的寻优模式[J].系统工程理论与实践,2002,22(3):76-82.
- [25] 江铭炎,袁东风.人工鱼群算法及其应用[M].北京:科学出版社,2012.
- [26] LELAND W E, TAQQU M S, WILLINGER W. On the self-similar nature of Ethernet traffic(extended version)[J]. IEEE/ACM Trans on Networking,1994,2(1):1-15.
- [27] PAXSON V, FLOYD S. Wide-area traffic: the failure of poisson modeling[J]. IEEE/ACM Trans on Networking,1996,3(3):226-244.
- [28] CLEGG R G. A practical guide to measuring the Hurst parameter[J]. International Journal of Simulation: Systems, Science & Technology,2006,7(2):3-4.
- [29] 郝清民. R/S 系列分析的非线性估计及应用[J].系统工程理论与实践,2005,25(3):80-85.

(上接第855页)估问题转换为分类问题,避免了指标权重确定的主观性,具有泛化能力强、建模简单且便于计算机实现等优点。在下一步的研究工作中,将对该模型中的隶属度函数进行改进,使得分类过程中相对距离的不确定性转换更为科学合理,评估结果更为可靠。

参考文献:

- [1] ELLISON R J, LINGER R C, LONGSTAFF T, et al. Survivable network system analysis: a case study[J]. IEEE Software, 1999, 16(4):70-77.
- [2] WANG H A, YAN Su, LIU Peng. A semi-Markov survivability evaluation model for intrusion tolerant database systems[C]//Proc of International Conference on Availability, Reliability and Security. 2010: 104-111.
- [3] 蔡均平,肖治庭,李雪冬.基于云模型的军事信息网络可生存性评估[J].武汉理工大学学报,2010,32(20):11-15.
- [4] 党德朋,孟真.基于支持向量机的信息安全风险评估[J].华中科技大学学报,2010,38(3):46-49.
- [5] 温祥西,孟相如.基于支持向量机和云模型的网络健康状态评估[J].北京邮电大学学报,2012,35(1):10-14.
- [6] 刘雨,陈进,潘玉娜,等.基于 SVDD 与信息融合技术的设备性能退化评估[J].振动与冲击,2009,28(9):21-24.
- [7] TAX D M J, DUIN R P W. Support vector domain description[J]. Pattern Recognition Letters,1999,20(11-13):1191-1199.
- [8] TAX D M J. Support vector data description[J]. Machine Learning, 2004,54(1):45-66.
- [9] 张庆,徐光华,王晶,等.基于支持向量域描述的多故障诊断动态模型[J].西安交通大学学报,2007,41(5):593-597.
- [10] BARZILAY O, BRAILOVSKY V L. On domain knowledge and feature selection using a support vector machines[J]. Pattern Recognition Letters,1999,20(5):475-484.
- [11] KENNEDY J, EBERHART R. Particle swarm optimization[C]//Proc of IEEE Conference on Neural Networks. Piscataway: IEEE Press, 1995:1942-1948.