

粒子群算法优化相空间重构参数的网络流量预测模型

曾伟¹, 黄亮²

(1. 华东交通大学 信息工程学院, 南昌 330013; 2. 上饶师范学院 数学与计算机科学学院, 江西 上饶 334001)

摘要: 在网络流量预测过程中, 相空间重构参数是影响预测性能的重要方面, 传统参数分开优化, 为了提高网络流量的预测精度, 提出一种粒子群算法优化相空间重构参数的网络流量预测模型 (PSO-BPNN); 首先将 BP 神经网络作为学习算法, 然后采用粒子群算法对相空间重构参数—延迟时间和嵌入维进行联合优化, 并重构网络流量序列, 最后以小波 BP 神经网络建立最优络流量预测模型, 并采用仿真实验对模型性能进行分析, 结果表明, PSO-BPNN 提高了网络流量的预测精度。

关键词: 网络流量; 相空间重构; 粒子群算法; 嵌入维; 延迟时间

Network Traffic Prediction Based on Phase Space Reconstruction Optimized By Particle Swarm Optimization Algorithm

Zeng Wei¹, Huang Liang²

(1. School of Information Engineering, East China JiaoTong University, Nanchang 330013, China;

2. Shangrao Normal University, School of Mathematics&Computer Science, Shangrao 334001, China)

Abstract: Parameters of phase space reconstruction are very important in network traffic prediction which is solved separately traditionally. In order to improve the prediction accuracy of network traffic, a novel network traffic prediction model (PSO-BPNN) is proposed in this paper based on particle swarm optimization algorithm and BP neural network. Firstly, BP neural network is taken as prediction algorithm, and the optimal delay time (τ) and embedding dimension (m) are obtained by particle swarm optimization and the network traffic series are reconstructed, finally, network traffic prediction models are established based on reconstruction the network traffic series, and simulation experiments are carried out to test the performance of network traffic prediction model. The results show that PSO-BPNN has improved the prediction accuracy of network traffic.

Keywords: network traffic; phase space reconstruction; particle swarm optimization algorithm; embedding dimension; delay time

0 引言

由于网络流量受到多因素影响, 传统线性模型难以建立准确的预测模型^[1]。基于支持向量机、神经网络的非线性预测方法能够较好地描述网络流量的非线性、非平稳性等变化特点, 预测精度高, 广泛应用于网络流量预测^[2-4]。近年来, 一些学者将混沌理论应用于网络流量建模与预测中, 其中相空间重构 (phase space reconstruction, PSR) 是基于混沌理论的网络流量预测的关键和基础^[5]。相空间重构通过两个关键参数—延迟时间 (τ) 和嵌入维 (m), 挖掘隐含于网络流量序列的信息, 因此 τ 和 m 直接影响网络流量的预测精度^[6]。对于 τ 和 m 优化问题, 目前主要有两类方法: 一类方法认为 τ 和 m 不相关, 即 τ 和 m 需要独立选择, 如采用平均位移法、自相关法、互信息法选择 τ ^[7]; 采用 Cao 法、关联维数法、假近邻法选择 m ^[8]。另一类方法认为 τ 和 m 相关, 应该进行组合优化, 主要有 C-C 法和时间窗口法等, 它们明确 τ 、 m 间有一定的乘积关系, 同时, 大量研究表明, τ 、 m 是密切相关, 对 τ 、 m 进行组合选择可以更加反映实际网络流量变化趋势^[9]。但是,

目前 τ 、 m 组合优化方法还在一些不足, 因此难以获得整体性能最优的网络流量预测模型。

为了提高网络流量的预测精度, 提出一种粒子群算法 (particle swarm optimization, PSO) 优化相空间重构参数的网络流量预测模型 (PSO-BPNN), 通过具体网络流量序列进行仿真实验和分析, 验证 PSO-BPNN 的网络流量预测性能。

1 小波 BP 神经网络和相空间重构

1.1 小波定义

小波分析是近年发展起来的一种用于信号分析的数学方法, 它源于傅里叶分析。傅里叶变换提供了频率域的信息, 但时间方面的局部化信息却基本丢失^[11]。而小波变换则可通过平移和伸缩变换处理获得信号的时频域局部化信息。设 $\psi(x)$ 为母小波, 通过缩放和平移可生成一组小波基函数 $\{\psi_{a,b}(x)\}$:

$$\psi_{a,b}(x) = |a|^{-1/2} \left(\frac{x-b}{a} \right) \quad (1)$$

其中: a, b 为缩放因子和平移参数。

以小波 $\psi(x)$ 为基的连续小波变换定义为:

$$W_f(a,b) = [f, \psi_{a,b}] = \int_{-\infty}^{+\infty} f(x) \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) dx \quad (2)$$

其中: 母小波函数 $\psi(x)$ 满足允许条件:

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\dot{\psi}(\omega)|}{|\omega|} d\omega < \infty \quad (3)$$

收稿日期: 2014-04-22; 修回日期: 2014-05-30。

基金项目: 江西省教育厅青年基金项目 (GJJ13704)。

作者简介: 曾伟 (1978-), 女, 江西樟树人, 讲师, 硕士, 主要从事网络技术、数据库技术、计算机应用软件开发方向的研究。

式中, $\hat{\psi}(\omega)$ 是 $\psi(x)$ 的傅里叶变换。

1.2 小波 BP 神经网络

BP 网络是误差反向传播网络, 隶属于前馈网络中的一种。由于其强大的非线性映射能力和自学习能力, BP 网络在众多领域中得到了广泛的应用。小波神经网络 (Wavelet Neural Network, WNN) 则有机的结合了小波分析和神经网络的优点, 加快了网络的收敛能力。设输入层、隐含层和输出层节点数分别为 I 、 H 、 O , 样本数为 P , 则隐含层第 j 个节点的输入为:

$$net_j = \sum_{i=1}^I \frac{(w_{ij}x_i - b_j)}{a_j} \quad (4)$$

输出层第 k 个节点的输出为:

$$O_k = w_{jk}f\left(\sum_{j=1}^H w_{jk}(\psi_{a,b}(net_j) + \delta_j)\right) + \delta_k \quad (5)$$

其中: δ_j 和 δ_k 是隐含层和输出层阈值。

小波 BP 神经网络在训练过程中, 计算输出值和真实值的误差函数, 当误差不满足精度要求时, 按梯度下降方向调节权值、阈值和小波参数减小误差, 直至满足精度要求。梯度为:

$$\frac{\partial E}{\partial w_{jk}} = \sum_{p=1}^P (d_k - O_k)O_k(1 - O_k)y_j \quad (6)$$

$$\frac{\partial E}{\partial w_{ij}} = \sum_{p=1}^P \left(\sum_{k=1}^O (d_k - O_k)O_k(1 - O_k)w_{jk}\right) \frac{\psi'_{a,b}(net_j)}{a_j} x_i \quad (7)$$

$$\frac{\partial E}{\partial \delta_k} = \sum_{p=1}^P (d_k - O_k)O_k(1 - O_k) \quad (8)$$

$$\frac{\partial E}{\partial \delta_j} = \sum_{p=1}^P \left(\sum_{k=1}^O (d_k - O_k)O_k(1 - O_k)w_{jk}\right) \frac{\psi'_{a,b}(net_j)}{a_j} \quad (9)$$

$$\frac{\partial E}{\partial b_j} = - \sum_{p=1}^P \left(\sum_{k=1}^O (d_k - O_k)O_k(1 - O_k)w_{jk}\right) \frac{\psi'_{a,b}(net_j)}{a_j} \quad (10)$$

$$\frac{\partial E}{\partial a_j} = - \sum_{p=1}^P \left(\sum_{k=1}^O (d_k - O_k)O_k(1 - O_k)w_{jk}\right) \frac{\psi'_{a,b}(net_j)}{a_j} net_j \quad (11)$$

带有附加动量因子的权值、阈值、尺度因子和水平因子修正量为:

$$\Delta w_{jk}(t+1) = (1 - mc)\eta \frac{\partial E}{\partial w_{jk}} + mc\Delta w_{jk}(t) \quad (12)$$

$$\Delta w_{ij}(t+1) = (1 - mc)\eta \frac{\partial E}{\partial w_{ij}} + mc\Delta w_{ij}(t) \quad (13)$$

$$\Delta \delta_k(t+1) = (1 - mc)\eta \frac{\partial E}{\partial \delta_k} + mc\Delta \delta_k(t) \quad (14)$$

$$\Delta \delta_j(t+1) = (1 - mc)\eta \frac{\partial E}{\partial \delta_j} + mc\Delta \delta_j(t) \quad (15)$$

$$\Delta b_j(t+1) = (1 - mc)\eta \frac{\partial E}{\partial b_j} + mc\Delta b_j(t) \quad (16)$$

$$\Delta a_j(t+1) = (1 - mc)\eta \frac{\partial E}{\partial a_j} + mc\Delta a_j(t) \quad (17)$$

式中, Δw 是权值增量, E 是误差函数, 常数 η 表示学习速率, t 为训练次数, mc 为动量因子, $0 < mc < 1$ 。

1.3 相空间重构

根据 Takens 定理可知, 只要选取合理的 τ 和 m , 相空间就可以重构与原始系统拓扑等价的动力系统^[12]。对于一个网络流量时间序列 $\{x(t)\}$, 选择适当的 τ 、 m 就得到一个 m 维的网络流量序列:

$$x(t) = [x(t), x(t - \tau), \dots, x(t - (m - 1)\tau)]^T \quad (18)$$

在重构的相空间中, 存在着光滑的非线性映射 $f: R^m \rightarrow R^1$, 即:

$$x(t + \tau) = f(x(t)) \quad (19)$$

通过小波 BP 神经网络对映射函数 $f(\cdot)$ 进行无限逼近, 建立一个精度高的网络流量预测模型, 该模型预测性能与 τ 、 m 密切相关。

1.4 PSO-BPNN 的流量预测

τ , m 联合优化显然是一个组合优化问题, 采用穷举法进行求解, 计算量很大, PSO 是一种基于种群迭代的随机搜索算法, 通过粒子之间的协作找到问题最优解。在粒子群算法 (PSO) 中, 每个粒子均具有自己的位置和速度, 粒子的位置代表问题的一个可能解, 每个粒子按式 (20) 和 (21) 更新自己的速度和位置:

$$v^{i+1}(i, k) = \omega v^i(i, k) + c_1 \text{rand}[p_{best}^i(i, k) - x^i(i, k)] + c_2 \text{rand}[g_{best}^i(i, k) - x^i(i, k)] \quad (20)$$

$$x^{i+1}(i, k) = x^i(i, k) + v^i(i, k) \quad (21)$$

式中, t 表示迭代次数; ω 称为惯性权重; c_1 、 c_2 为学习因子, rand 是在 $0 \sim 1$ 之间的随机函数; p_{best} 表示粒子本身经历过的最好位置; g_{best} 表示种群经历过的历史最好位置。

适应度函数是根据目标函数确定的用于区分群体中粒子好坏的标准, 是算法演化过程的驱动力, 本文采用网络流量的预测精度作为粒子的适应度函数, 即:

$$\text{fitness}(i) = \sum_{j=1}^I (y_j - \bar{y}_j)^2 \quad (22)$$

式中, y_j 和 \bar{y}_j 分别表示模型预测值和网络流量实际值。

基于 PSO-BPNN 网络流量预测模型的工作流程如图 1 的所示。

2 仿真实验

2.1 数据集来源

在 Pentium Dual-Core 3.80 GHz CPU, 4 GB RAM, Windows XP, 所有程序采用 VC++6.0 编程实现。数据来源于某企业网络中心服务器 2012 年 11 月 1 日至 12 月 25 日的每小时网络流量, 收集到 1 000 个数据点, 网络流量变化如图 2 所示。选择前 800 个数据组成训练集, 建立预测模型; 其余 200 个数据组成预测集, 以检验模型的有效性。采用相空间重构参数分开、独立优化的网络流量预测模型进行对比实验。

2.2 数据预处理

为了加快小波 BP 神经网络的网络流量建模效率, 对网络流量数据进行归一化处理, 具体为:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (23)$$

式中, x_{\max} 和 x_{\min} 分别表示最大值和最小值。

2.3 结果与分析

2.3.1 参数分开优化的预测结果

互信息函数图如图 3 所示。从图 3 可知, 网络流量时间序列的最佳时间延迟为 $\tau = 4$ 。不同嵌入维数下的关联维数变化曲线如图 4 所示。从图 4 可知, 网络流量最优 $m = 5$ 。

采用最佳参数 $\tau = 4$, $m = 5$ 对网络流量序列进行重构, 然后将重构后的训练集输入小波 BP 神经网络进行训练, 并对测试

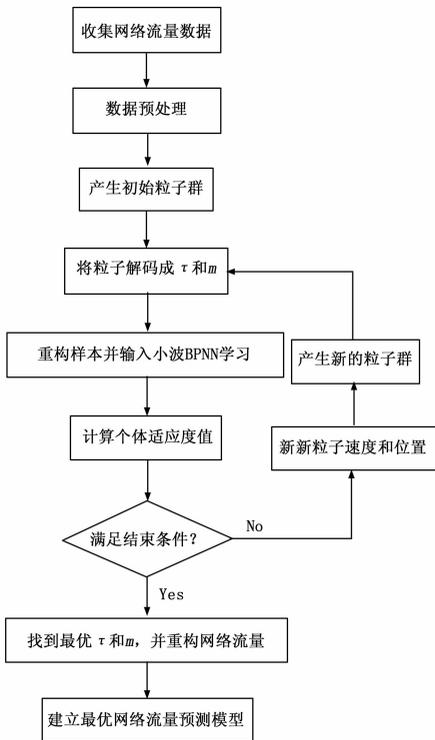


图 1 网络流量的建模流程

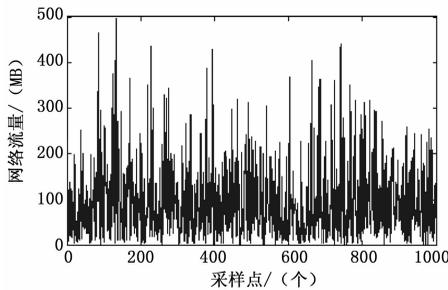


图 2 网络流量数据

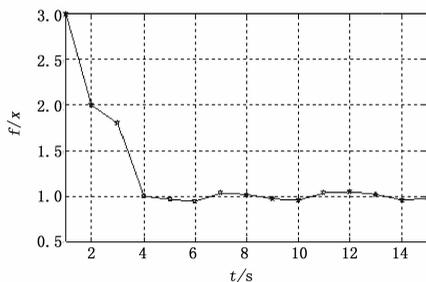


图 3 时间延迟求解

集进行预测, 结果如图 5 所示。对图 5 进行分析可知, 相空间参数分开、单独优化的网络流量的预测值与实际值偏差较大, 预测误差大, 预测精度低, 不能满足网络流量预测的实际要求。

2.3.2 组合优化参数 (PSO-BPNN) 的网络流量预测性能分析
采用 PSO 对 τ 和 m 进行组合优化, 得到最优 $(\tau, m) =$

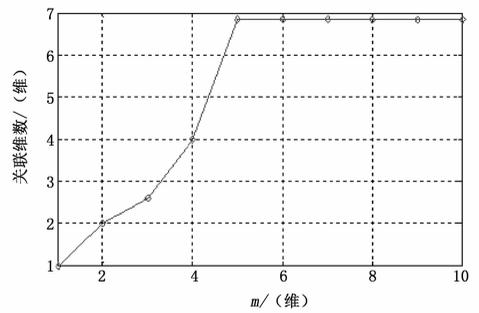


图 4 嵌入维数的计算

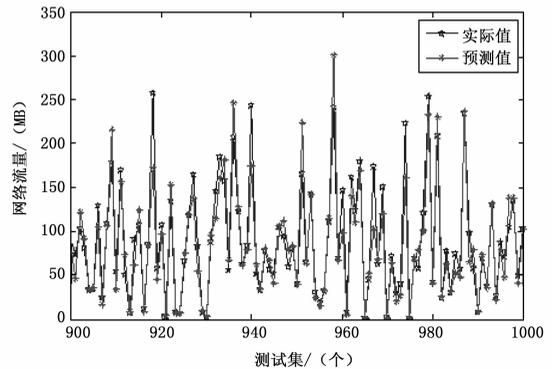


图 5 相空间重构参数分开优化模型的预测结果

(2, 6), 采用 $(\tau, m) = (2, 6)$ 对网络流量进行重构, 并将训练集输入小波 BP 神经网络建立最优的网络流预测模型, 预测结果如图 6 所示。从图 6 可知, PSO-BPNN 用 PSO 对相空间重构参数进行联合优化, 充分挖掘了隐含于网络流量序列中的信息, 建立的模型可以准确刻画网络流量的演化过程, 相对于对比算法, PSO-BPNN 提高了网络流量预测精度, 是一种有效的网络流量预测方法。

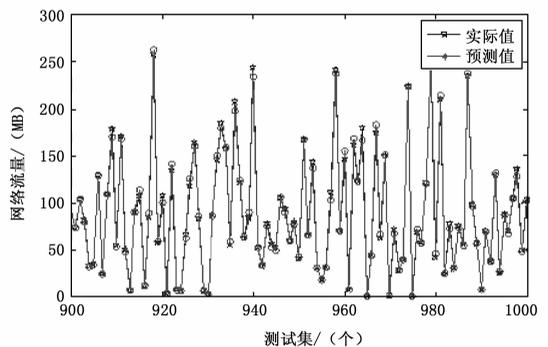


图 6 PSO-BPNN 的预测结果

2.4 与其它网络流量预测模型性能对比

采用多元线性回归 (MLR)、时间序列预测模型 (ARIMA) 和支持向量机 (SVM)、相关向量机 (RVM) 进行对比实验, 并采用 RMSE、MAPE 作为模型性能评价标准, RMSE、MAPE 定义如下:

(下转第 3043 页)

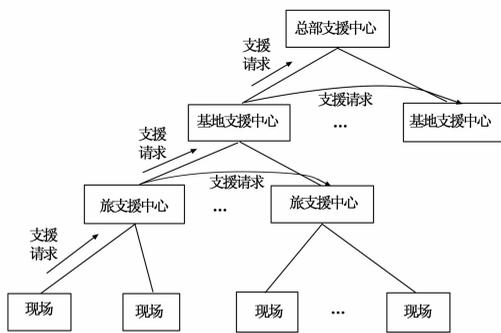


图 4 远程技术支援系统体系结构

视频信息的同步。

系统分别在各旅指挥所部署旅级视频服务器,在基地部署基地视频服务器,在总部部署总部视频服务器。这种情况下,即使旅和基地的网络没有连接,现场的音视频信息可以在相连的各个旅指挥所之间进行同步,不会影响日常的工作。在特殊时期,当旅指挥所、基地指挥所和总部指挥所网络连通时,现场的音视频信息同样可以采用对等有向的同步方式,在各级各

节点之间进行信息同步。远程技术支援系统中音视频数据同步手段采用狭义的同步,数据同步的对象是关系数据库的同步。

5 结束语

建立多级支援中心的远程技术支援系统,不仅适应目前部队的网络建设环境和工作模式,而且能够实现各种信息的有效快速传递,使得支援过程更加及时便捷,提高了装备的保障能力。对等有向的数据同步方式改变了以往分布式系统必须建立一个服务的模式,使得每一个节点既可以变成了服务端,也可以变成客户端,提高了分布式系统的数据同步效率。

参考文献:

[1] 孙 洁, 于 涵. 基于多中心支持的远程技术支援体系 [A]. 第二十二届测试与故障诊断技术研讨会 [C]. 2013.
 [2] 蒲志林. 网络环境下分布式数据库同步机制的研究 [D]. 大连: 大连海事大学, 2003.
 [3] 吴宇翔. G/S 模式下分布式空间数据服务器群的数据同步研究 [D]. 成都: 成都理工大学, 2010.

(上接第 3016 页)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (24)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|^2} \quad (25)$$

式中, n 表示测试集的样本数; y_i 和 \hat{y}_i 分别为网络流量的实际值和预测值。

PSO-BPNN、MLR、ARIMA、SVM、RVM 各模型的预测结果的 RMSE 和 MAPE 见表 1。从表 1 可知,相对于 MLR、ARIMA、SVM、RVM 等模型的预测结果,PSO-BPNN 的预测误差更小,具有更明显的优势,获得了更加理想的网络流量预测效果。

表 1 PSO-BPNN 与当前经典网络流量模型的性对比

预测模型	RMSE	MAPE/%
MLR	6.402	5.392
ARIMA	6.228	4.664
SVM	5.796	3.655
RVM	5.226	4.584
PSO-BPNN	3.411	2.962

3 结语

为了提高网络流量预测精度,针对相空间重构参数优化问题,提出一种基于 PSO-BPNN 的网络流量预测模型,并通过仿真实验对模型性能进行测试。结果表明,PSO-BPNN 克服了传统方法存在的不足,提高了网络流量的预测精度,建立的模型可以准确刻画网络流量的演化过程,为具有混沌性的网络流量提供了一种新的建模方法。

参考文献:

[1] 巩林明, 张振国. 基于灰色小波的网络流量组合预测模型 [J]. 计算机工程与设计, 2010, 31 (8): 1660-1661.
 [2] 田妮莉, 喻 莉. 一种基于小波变换和 FIR 神经网络的广域网流量预测模型 [J]. 电子与信息学报, 2008, 30 (10): 2549-2555.
 [3] 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.
 [4] 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): 130-131.
 [5] 刘 渊, 王 鹏. 融合小波变换与贝叶斯 LS-SVM 的网络流量预测 [J]. 计算机应用研究, 2009, 26 (6): 2229-2231.
 [6] 罗赞骞, 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 [J]. 计算机科学, 2009, 6 (7): 244-246.
 [7] 姜 明, 吴春明, 张 曼, 等. 网络流量预测中的时间序列模型比较研究 [J]. 电子学报, 2009, 37 (11): 2353-2358.
 [8] Silva C G. Time series forecasting with a nonlinear model and the scatter search meta-heuristic [J]. Information Sciences, 2008, 178 (16): 3288-3299.
 [9] 融合提升小波降噪和 LSSVM 的网络流量在线预测 [J]. 计算机应用, 2012, 32 (2): 340-342.
 [10] Qi H L, Zhao H, Liu W W, et al. Parameters optimization and nonlinearity analysis of grating eddy current displacement sensor using neural network and genetic algorithm [J]. Journal of Zhejiang University Science A, 2009, 10 (8): 1205-1212.
 [11] 刘 绚, 刘天琪. 基于小波变换和遗传算法优化神经网络负荷预测 [J]. 四川电力技术, 2010, 33 (3): 15-19.
 [12] Ames T, Rego C Glover F. Multistart. Babu search and diversification strategies for the quadratic assignment problem [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part A Systems and Humans, 2009, 39: 579-596.