

遗传算法优化小波神经网络的网络流量预测^①

王雪松¹, 赵跃龙²

¹(佛山职业技术学院 电子信息系, 佛山 528137)

²(华南理工大学 计算机科学与工程学院, 广州 510640)

摘要: 为了提高网络流量的预测精度, 克服小波神经网络收敛速度慢、易陷入局部最优的缺点, 提出一种遗传算法优化小波神经网络的网络流量预测模型。首先计算延迟时间和嵌入维数, 构建小波神经网络的学习样本, 然后采用小波神经网络对网络流训练集进行学习, 并采用改进遗传算法对小波神经网络参数进行全局寻优, 提高收敛速度和网络学习精度, 最后采用网络流量数据对模型性能进行仿真分析。结果表明, 相对于对比模型, 本文模型的平均误差大幅度降低, 训练次数急剧减, 减小了二次优化训练的次数, 具有更大的实际应用价值。

关键词: 网络流量; 小波神经网络; 遗传算法; 参数优化

Network Traffic Prediction Based on Wavelet Neural Network and Genetic Algorithm

WANG Xue-Song¹, ZHAO Yue-Long²

¹(Department of Electronic Information, Foshan Polytechnic College, Foshan 528137, China)

²(School of Computer and Engineer, South China University of Technology, Guangzhou 510640, China)

Abstract: In order to overcome the shortcomings of wavelet neural network and improve prediction precision of network traffic, a novel network traffic prediction model is proposed based on the wavelet neural network and genetic algorithm in this paper. Firstly, the time delay and embedding dimension of network traffic are calculated to construct the learning samples of wavelet neural network. Then training samples are input to wavelet neural network to learn in which improved genetic algorithm is used to optimize the parameters of wavelet neural network. Finally, the performance of model is tested by simulation experiment using network traffic data. The results show that the proposed model has reduced the prediction error and the number of training has reduced sharply compared with other model, so it has great practical application value.

Key words: network traffic; wavelet neural network; genetic algorithm; parameter optimization

随着网络应用范围日益扩大, 网络拥塞频繁, 网络流量的预测结果可以为网络管理员提供技术帮助, 对网络实施相应的控制策略, 提高网络服务质量, 因此网络流量预测是一直网络研究中的重点课题^[1]。

目前, 国内外学者对网络流量预测问题进行大量研究, 传统预测模型有多元线性回归法、滑动平均法、差分自回归移动平均法、指数平滑法等^[2,3], 它们均基于线性理论进行建模, 对随机性、时变性等的网络流量难以达到准确对预测, 预测精度有待提高^[4]。针对网

络流量变化特点, 一些学者将非线性理论引入到网络流量预测中, 出现支持向量机、灰色理论、高斯过程回归、神经网络等预测模型^[5-7], 它们较好描述了网络流量变化趋势, 网络流量的预测精度大幅度提高^[8]。由于网络流量具有多尺度特性, 有学者采用小波神经网络(WNN)对网络流量进行建模, 取得了不错的预测效果^[9]; 然而小波神经网络性能与网络参数密切相关, 当前主要采用梯度下降法对网络参数进行优化, 梯度变化方向相对固定, 限制了小波神经网络参数优化的

① 基金项目: 国家自然科学基金(60573145); 广东省教育厅项目(2010tj446)

收稿时间: 2014-05-06; 收到修改稿时间: 2014-06-03

方向,难以适应网络流量随机性和时变等变化特点,且存在易陷入局部极小点和引起振荡效应等不足^[10].

为了克服小波神经网络小波神经网络收敛速度慢、易陷入局部最优的缺点以提高网络流量预测精度,提出一种遗传算法优化小波神经网络的网络流量预测模型(GA-WNN),并采用仿真实例对模型性能进行测试与分析.仿真结果表明,相于其它模型,GA-WNN的各项预测误差指标更小,提高了网络流量的预测精度.

1 小波神经网络模型

小波神经网络(Wavelet Neural Network, WNN)是一种结合小波变换和神经网络的新型前馈网络,激活函数采用非线性小波基函数代替 Sigmoid 函数作,大量研究表明,一个非线性映射问题可以通过 3 层前馈网络以任意的精度来逼近^[11].假设输入层共有 m 个输入节点数, L 个隐含层节点数,只有 1 个输出层节点数,小波神经网络网络结构如图 1 所示.

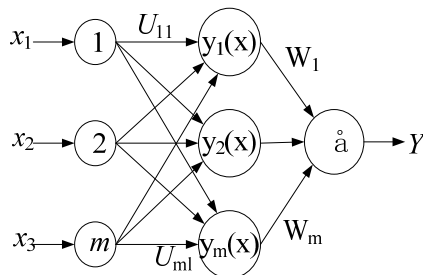


图 1 小波神经网络结构图

\hat{Y} 表示网络的输出期望值,其计算公式为

$$\hat{Y} = \sum_{i=1}^L W_i \psi \left(\frac{x - b_i}{a_i} \right) \quad (1)$$

式中, W_i 是第 i 个隐含层节点输出的权重值.

令 t_i 表示第 i 个隐含层节点的输入,则

$$t_i = \sum_{j=1}^m U_{ij} x_j \quad (2)$$

式中, U_{ij} 表示第 i 个小波基单元与第 j 个输入值间的权重.

由式(1)和(2)可以得到

$$\hat{Y} = \sum_{i=1}^L W_i \psi \left(\frac{\sum_{j=1}^m U_{ij} x_j - b_i}{a_i} \right) \quad (3)$$

假设样本数据为

$(x_k, y_k) \quad x_k \in R^m, y_k \in R, k=1, 2, \dots, N$, x_k 和 y_k 分别表示网络的输入和输出,根据训练集对采用梯度下降法对小波神经网络网络进行训练,确定参数 U_{ij} , W_i , a_i , b_i , 并且使误差能量函数最小

$$E = \frac{1}{2} \sum_{k=1}^N (y_k - \hat{y}_k)^2 \quad (4)$$

式中, \hat{y}_k 表示第 k 个输入样本的网络输出.

隐含层神经元函数采用 Morlet 小波函数

$$\psi(x) = \cos(cx) * e^{-\frac{x^2}{2}} \quad (5)$$

式中, c 为一常量.

对于输入 x_k , 相应的网络输出为

$$\hat{y}_k = \sum_{i=1}^L W_i \psi \left(\frac{\sum_{j=1}^m U_{ij} x_{kj} - b_i}{a_i} \right) \quad (6)$$

将式(6)代入式(4)可得

$$E = \frac{1}{2} \sum_{k=1}^N \left(y_k - \sum_{i=1}^L W_i \psi \left(\frac{\sum_{j=1}^m U_{ij} x_{kj} - b_i}{a_i} \right) \right)^2 \quad (7)$$

2 改进遗传算法的小波神经网络预测模型

2.1 改进遗传算法

遗传算法是一种自适应随机优化搜索算法,具有隐含并行性和全局寻优能力特点,能很好的克服梯度下降法的缺点^[13].针对小波神经参数优化的具体问题,本文对遗传操作算子进行了相应的,具体如下:

1) 个体适应值的计算和调整方式为

$$\begin{cases} e = \frac{1}{2} (\hat{y}_k(x_p) - y_k(x_p))^2 \\ f_i = \frac{1}{e} \\ f' = af + b \end{cases} \quad (8)$$

式中, f 和 f' 分别表示调整前和调后的个体适应度值; a, b 为调整的系数.

为了避免调整后出现负数的适应度值,再对 a, b 的值进行调整,如果 $f_{\min} > (cf_{\text{avg}} - f_{\max}) / (c - 1.0)$, 采用式(9)进行调整,否则采用式(10)进行调整.

$$\begin{cases} a = (c - 1.0) f_{\text{avg}} / \Delta_1 \\ b = f_{\text{avg}} (f_{\max} - cf_{\text{avg}}) / \Delta_1 \end{cases} \quad (9)$$

$$\begin{cases} a = f_{avg} / \Delta_2 \\ b = (-f_{min} f_{avg}) / \Delta_2 \end{cases} \quad (10)$$

式中, $\Delta_1 = f_{max} - f_{avg}$; $\Delta_2 = f_{avg} - f_{min}$; $c = 2.0$; f_{min} 为种群最小值; f_{max} 种群最大值; f_{avg} 为平均适应值。

2) 选择操作. 采用适应度值比例选择方法, 适应度值越大, 选择概率越大. 选择概率为:

$$p_s = p \times f_i / \sum f_i \quad (11)$$

式中, f_i 为适应度值, p 为种群数目。

3) 交叉操作. 根据交叉概率 p_c 从种群中选择两个个体进行交叉, 交换它们部分基因, 得到新的个体, 增加搜索范围. 具体交叉方式为

$$P_c = \begin{cases} k_1(f_{max} - f'_c) / (f_{max} - f_{avg}) & f'_c \geq f_{avg} \\ k_2 & otherwise \end{cases} \quad (12)$$

式中 $k_1 = k_2 = 1.0$, f'_c 为两个父代中待交叉的比较大的适应值。

(4) 变异操作. 根据概率 P_m 从相应的补集中选出某个值替换某个个体中的每一位, 具体变异方式为

$$P_m(g) = 0.001 + NG \times Cof \quad (13)$$

式中, g 为遗传代数; NG 为未进化的代数; Cof 为系数。

为了验证本文对遗传算法改进的有效性, 选用 1 常用基准函数遗传算法改进前后性能进行测试, 测试函数如下:

$$f_1 = \frac{1}{4000}(x_1^2 + x_2^2) - \cos x_1 \cos(\frac{x_2}{\sqrt{2}}) + 1 \quad (14)$$

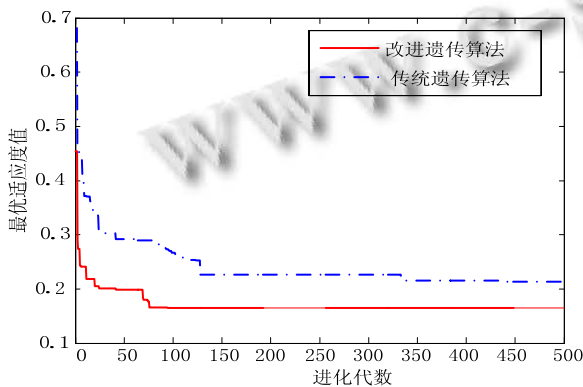


图 2 遗传算改进前后的收敛性对比

遗传算法改进前后的收敛速度和精度如图 2 所示。

从图 2 可知, 改进后遗传算法的收敛速度明显加快, 较好的解决了传统遗传算法的缺点, 这样更快的找到小波神经网络参数, 减少训练次数, 提高网络流量建模效率。

2.2 GA-WANN 的工作步骤

GA-WNN 利用遗传算法良好的全局搜索能力对神经网络权值进行优化上, 解决网络流量训练过程易陷入局部极小的难题, 同时利用神经网络的强大函数映射及逼近能力, 对网络流量的复杂变化特点进行建模, 以获得比较理想的预测结果, 具体工作步骤如下:

- 1) 网络参数的初始化, 确定 a_i 、 b_i 、 U_{ij} 、 W_i 的初始值, 并对遗传算法的初始参数进行设定, 每个染色体代表一种网络结构。
- 2) 根据式(14)计算个体 n 的适应度函数 $f(n)$, 并且根据每一个个体的适应值, 采用轮盘赌选择个体进行下一代的概率。
- 3) 按交叉概率 P_c 从种群中随机选择两个个体进行交叉, 并且根据式(13) 交叉概率 P_c 。
- 4) 利用概率 P_m 变异产生新的个体, 变异方式为式(14)。
- 5) 将由 P_m 变异产生的新个体插入到种群 P 中, 计算新个体的适应值。
- 6) 判断是否达到预定值, 是则结束, 否则转步骤 3)。
- 7) 将最终群体中的最优个体解码作为优化的 WNN 连接权和伸缩平移因子。
- 8) 采用最优参数建立网络流量预测模型。

3 仿真实验

3.1 数据来源

数据来源于 <http://newsfeed.ntcu.net/~news/2012/> 的每小时网络流量, 其获得 720 样本 $\{x(i)\}$, $i=1,2,\dots,720$, 具体如图 3 所示. 选择前 620 个样本作为训练集, 对 WNN 的参数进行训练, 建立 GA-WNN 的网络流量预测模型, 其余 100 个样本作为测试集对模型性能进行检验. 所有实验在 P4 4 核 3.0 CPU, 4G RAM, Windows XP 的计算机上, 采用 Matlab 2012 编程实现。

由于网络流量数值变化较大时, 对 WNN 学习产生不利影响, 因此建模之前对网络流量进行归一化处理, 具体为:

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (15)$$

式中, x_{\max} 和 x_{\min} 为最大值和最小值; x_i 为原始值.

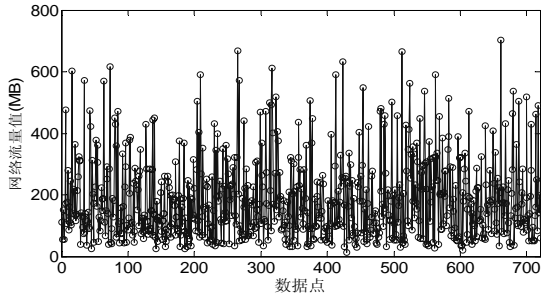
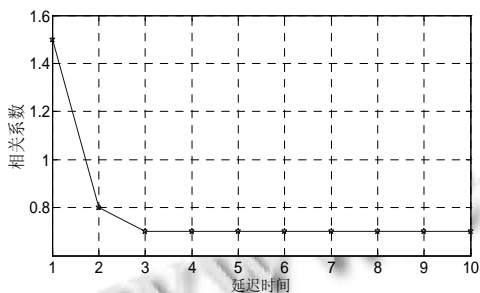


图 3 网络流量数据

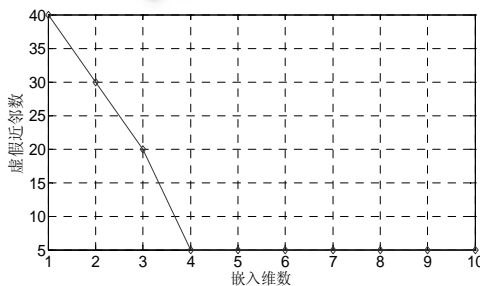
3.2 学习样本的构造

网络流量的自相关函数变化曲线如图 4(a)所示, 从图 4(a)可知, 最佳延迟时间 $\tau=3$. 虚假近邻数和嵌入维数间的变化曲线如图 4(b)所示, 从图 4(b)可知, 最优嵌入维数 $m=4$, 采用 $\tau=3, m=4$ 对网络流量数据重构, 得到 GA-WNN 的学习样本, 由于嵌入维数 $m=4$, 表示当前预测点, 与 4 个历史样本点密切相关, 因此仿真实验中的输入向量是多少维为 4.

综合上述可知, WNN 输入层神经元的数目为 4, 隐含层神经元个数为 4, 输出层神经元个数为 1, 因此 WNN 的网络结构为 4-4-1.



(a) 计算延迟时间



(b) 计算嵌入维数

图 4 延迟时间和嵌入维数的确定

3.3 对比模型及评价标准

为了使 GA-WNN 的预测结果具有可比性, 传统小波神经网络(WNN)进行对比实验, WNN 的参数为: 小波函数选择 Morlet 小波, 训练次数为 100, 动量因子为 0.01, 小波个数为 14. 遗传算法的交叉概率 $P_c=0.75$, 变异概率 $P_m=0.02$, 最大进化代数 $T_{\max}=500$. 由于待优化参数为 a_i, b_i, U_{ij}, W_i , 因此遗传算法的个体维数为 4, a_i, b_i, U_{ij}, W_i . 为了评价模型的预测精度, 选取平均绝对误差(MAE)、均方根误差(RMSE)和均方百分比误差(MSP)作为评价依据, 它们分别定义如下:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_{real} - Y_{pre}| \quad (16)$$

$$MSE = \frac{1}{n} \sqrt{\sum_{i=1}^n (Y_{real} - Y_{pre})^2} \quad (17)$$

$$MSP = \frac{1}{n} \sqrt{\sum_{i=1}^n \left(\frac{Y_{real} - Y_{pre}}{Y_{real}} \right)^2} \quad (18)$$

3.4 结果与分析

3.4.1 与 WNN 的预测结果比较

种群规模设置 10~30, 最后得到当种群数量的大小为 20 时, 效果好, 此时, WNN 和 GA-WNN 预测结果如图 5 所示. 从图 5 可以看出, 相对于 WNN, GA-WNN 网络流量的预测值和实际值变化曲线拟合性更好, 预测精度大幅度提高, 这主要由于采用遗传算法对 WNN 的参数进行优化, 较好的克服了梯度下降法参数初始值敏感, 易陷入局部极小和产生振荡效应, 获得全局最优的网络参数, 建立预测性能优异的网络流量预测模型.

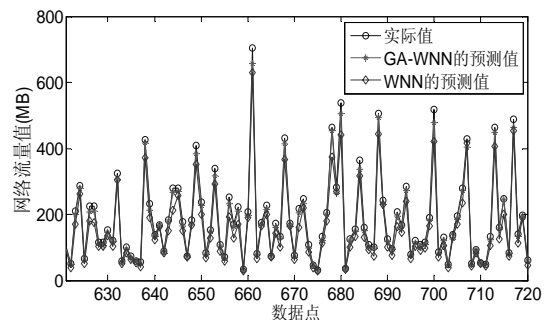


图 5 WNN 和 GA-WNN 的预测值和实际值对比

WNN 和 GA-WNN 的综合评价指标见表 1. 对表 1 进行分析, 可知, GA-WNN 的各项预测误差指标均小

于WNN,平均误差大大减少,网络流量预测精度明显提高,同时在GA-WNN的网络流量建模过程中,训练次数有原来的450次变为180次,大大减小了二次优化训练的次数,减小了WNN计算复杂度,提高了其收敛速度,可以较好的满足网络流量在线预测要求。

表1 WNN和GA-WNN的综合性能对比

评价标准	WNN	GA-WNN
MAE	20.752	14.764
RMSE	4.220	2.359
MSP	0.023	0.015

3.4.2 对含噪声网络流量的预测性能

为了测试GA-WNN模型的鲁棒性,采用一个含有噪声网络流量进行仿真实验,以测试模型的鲁棒性,含噪的网络流量数据如图6所示。对含噪的网络流量数据进行建模与预测,GA-WNN的网络流量预测结果如图7所示。从图7可知,GA-WNN获得了较理想的预测结果,这表明GA-WNN具有较强稳健性和鲁棒性,具有一定的抗干扰性能。

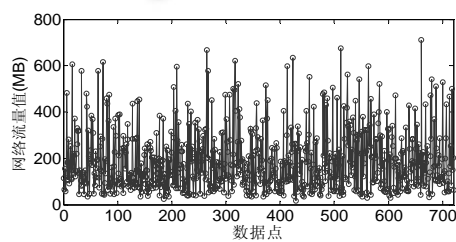


图6 含噪声的网络流量数据

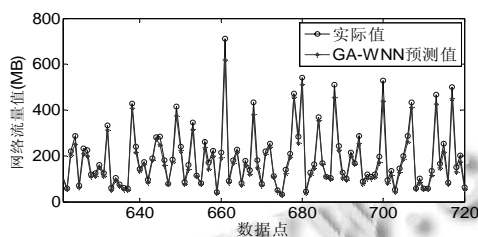


图7 GA-WNN的含噪声网络流量预测结果

4 结语

针对网络流量的非线性和时变性的特点,利用遗传算法的全局优化搜索能力和小波神经网络强大的非

线性预测能力,提出一种GA-WNN的网络流量预测模型,仿真结果表明,GA-WNN加快了网络的收敛速度,提高了网络流量的预测精度,为网络流量预测提供了一种新的有效途径。

参考文献

- 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):1609-1617.
- 党小超,郝占军.季节周期性Elman网络的网络流量分析与应用.计算机工程与应用,2010,46(28):98-101.
- 罗贲,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.
- 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.
- 王俊松,高志伟.基于RBF神经网络的网络流量建模及预测.计算机工程与应用,2008,44(13):6-7.
- 孙韩林,金跃辉,崔毅东,程时端.粗粒度网络流量的灰色模型预测.北京邮电大学学报,2010,33(1):71-75.
- 熊南,刘百芬.基于自适应粒子群优化LSSVM的网络流量在线预测.计算机应用与软件,2013,30(9):21-24,127.
- 周晓蕾,王万良,陈伟杰.基于小波变换和优化的SVM的网络流量预测模型.计算机应用与软件,2011,28(2):34-36,59.
- 雷霆,余镇危.一种网络流量预测的小波神经网络模型.计算机应用,2006,26(3):526-528.
- 于繁华,刘仁云,赵宏伟,等.改进的小波神经网络模型及应用.吉林大学学报(信息科学版),2005,23(5):489-492.
- 余健,郭平.基于改进小波神经网络的网络流量预测研究.计算机应用,2007,27(12):2986-2990.
- 李目,何怡刚,周少武,刘祖润.一种差分进化算法优化小波神经网络及其在弱信号检测中的应用.计算机应用与软件,2010,27(3):29-31,39.
- 刘靖洁,陈桂明,刘小方,杨庆.基于遗传算法的SVM参数组合优化.计算机应用与软件,2012,29(4):94-96,100.