

## 遗传优化神经网络的网络流量预测

赵清艳

(中山职业技术学院, 广东 中山 528404)

**摘 要:** 为了提高网络流量预测精度, 利用相空间重构和预测模型参数间的相互联系, 提出一种遗传算法优化神经网络的网络流量预测模型. 首先将相空间重构和神经网络参数进行编码, 网络流量预测精度作为目标函数, 然后通过遗传算法选择模型最优参数, 最后进行网络流量仿真实验. 实验结果表明相对传统预测模型, 遗传优化神经网络模型具有更高预测精度及稳定性更好.

**关键词:** 网络流量; 神经网络; 遗传算法; 相空间重构

**中图分类号:** TP393

**文献标识码:** A

**文章编号:** 1000-7180(2013)03-0132-04

## Network Traffic Prediction Based on Neural Network Optimized by GA

ZHAO Qing-yan

(Zhongshan Polytechnic, Zhongshan 528404, China)

**Abstract:** In order to improve the network traffic prediction accuracy, this paper proposes a network traffic prediction method based on RBF neural network optimized by genetic algorithm which uses the relation between phase space reconstruction and parameters of prediction model. Firstly, phase space reconstruction and the parameters of RBF neural network are coded, and then the model prediction accuracy is used as the objection function, and optimal parameters of the model are selected by genetic algorithm, lastly, the simulation experiments are carried out to test model's performance. The results show that, compared with the traditional models, the proposed model improves the prediction accuracy.

**Key words:** network traffic; neural network; genetic algorithm; phase space reconstruction

### 1 引言

如何提高网络流量预测精度是网络控制中的热点<sup>[1]</sup>. 针对网络流量预测问题, 学者对其进行了深入研究, 传统预测方法有: 回归分析、灰色模型和时间序列等<sup>[2-4]</sup>, 这些方法假设网络流量线性变化, 不能准确刻画网络流量的复杂非线性变化规律, 预测结果不理想<sup>[5]</sup>. 近年来, 随着非线性理论发展, 出现基于混沌理论的网络流量预测算法, 其包括相空间重构和模型参数优化两个问题. 相空间重构用于对隐藏于网络流量数据间的信息进行深入挖掘, 对原动力系统描述; 预测模型对网络流量数据进行学习和建模. RBF 神经网络是一种自学习能力强, 且

能够逼近任意非线性系统的机器学习算法, 克服了传统机器学习算法过拟合、局部最优缺陷, 成为网络流量的主要研究算法<sup>[6]</sup>. 在基于 RBF 神经网络的网络流量预测过程中, 相空间重构与 RBF 神经网络参数共同决定了网络流量的预测精度, 然而传统上两者常分开优化, 忽略两者间联系, 难以获得整体预测性能最优的网络流量预测模型. 为了使网络流量预测模型性能达到最佳, 必须对相空间重构与 RBF 神经网络参数必须同步优化<sup>[7]</sup>.

针对当前网络流量预测模型参数分开优化存在的缺陷, 提出一种遗传算法 (genetic algorithm, GA) 优化 RBF 神经网络的网络流量预测模型 (GA-RBF). 采用 GA 对网络流量预测模型参数进行同步

优化,并通过仿真实验验证模型的有效性.

## 2 相空间重构和 RBF 神经网络

### 2.1 相空间重构

相空间重构结果的优劣由最佳延迟时间( $\tau$ )和嵌入维( $m$ )决定,而相空间重构质量直接影响着模型的预测结果.根据 Takens 定理,对于没有噪声、无限的网络流量时间序列,与  $m$  没有联系,然而,在实际应用中,网络流量数据长度不能无限长且含有噪音,因此,大多数学者认为  $\tau$  和  $m$  应该同步优化.当前  $\tau$  和  $m$  联合优化方法主要有时间窗口法、C-C 法(coupled cluster method)和自相关联函数法等<sup>[8]</sup>,这些方法性能各异,判断标准具有较强主观性.

### 2.2 RBF 神经网络

RBF 神经网络是一种三层前馈式神经网络,其结构简单,收敛速度快,非线性预测能力强<sup>[9]</sup>.一个 RBF 神经网络由  $n$  个输入节点、 $m$  个隐层节点和 1 个输出节点组成,隐层节点 RBF 函数表示为:

$$h = \exp\left[-\frac{\|x - c\|^2}{2\sigma^2}\right] \quad (1)$$

式中  $c$  为第  $i$  个 RBF 隐节点的中心,  $w$  是输出权重,  $\sigma$  为 RBF 隐节点的宽度.当前 RBF 神经网络参数  $w, c, \sigma$  优化方法主要有梯度下降算法、遗传算法(GA)、粒子群优化算法(Particle Swarm Optimization, PSO)等.这些算法有的搜索速度慢,运算量大.有的没有考虑 RBF 神经网络参数和相空间重构间之间联系,难以获得模型整体预测性能最优的参数.

## 3 基于 GA-RBF 的网络流量预测模型

### 3.1 算法设计

#### (1) 染色体编码

每一个染色体代表一个参数组合,采用二进制编码方法对染色体进行编码,每一个染色体包括四部分: $\tau, m, w, c, \sigma$ ,具体如图 1 所示.

|        |     |     |     |          |
|--------|-----|-----|-----|----------|
| $\tau$ | $m$ | $w$ | $c$ | $\sigma$ |
|--------|-----|-----|-----|----------|

图 1 染色体编码设计

$$p = \min p + \frac{\max p - \min p}{2^l - 1} \times d \quad (2)$$

式中  $\min p$  和  $\max p$  分别为参数的最小值和最大值,  $l$  为位串长度,  $d$  为二进制位串.

#### (2) 适应度函数设计

适应度函数用来衡量一个个体的优劣,网络流

量预测模型参数优化的目标是提高网络流量的预测精度,因此将个体适应度函数  $f(x)$  定义:

$$f(x) = \frac{1}{\text{MSE}(\tau, m, w, c, \sigma)} \quad (3)$$

式中,  $\text{MSE}$ (mean squared error) 表示均方误差,定义如下:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4)$$

式中  $n$  表示网络流量训练集数目,  $y_i$  和  $\hat{y}_i$  分别表示网络流量期望值和预测值.

#### (3) 遗传操作

假定个体  $x_i$  的适应度为  $f(x_i)$ ,则该个体被选择的概率  $p_i$  为:

$$p_i = \frac{f(x_i)}{\sum f(x_j)} \quad (5)$$

式中  $\sum f(x_j)$  表示群体中的个体适应度的总和.根据变异概率选择一个染色体,然后随机选择一个变异点,采用  $0 \rightarrow 1, 1 \rightarrow 0$  的规则进行变异,变异后如果其值不在参数取值范围内,则重新变异.

### 3.2 模型参数优化过程

(1) 随机方式产生遗传算法规模为  $N_{\text{pop}}$  的初始种群  $R_{\text{gen}}$ .进化次数的初值为  $g = 1$ .

(2) 对每一个染色体进行反编码,根据  $\tau$  和  $m$  进行重相空间重构, RBF 神经网络根据  $w, c, \sigma$  进行建模,得到每一组参数的网络流量预测精度,通过预测结果的  $\text{MSE}$  对个体适应度值进行计算.

(3) 如果达到最大进化代数或者种群的最优适应度函数值多代不发生较大变化,则模型参数达到最优,跳转到步骤(5).

(4) 进化代数  $g = g + 1$ .根据选择、交叉和变异算子对种群进行更新,并跳转到步骤(2).

(5) 得到网络流量预测模型的最优参数  $\tau, m, w, c, \sigma$ .采用  $\tau, m$  对网络流量时间序列进行重构,  $w, c, \sigma$  作为 RBF 神经网络的参数进行训练建模,对测试集进行预测,输出模型对网络流量预测的结果.

## 4 仿真实验

### 4.1 数据集来源

为了验证 GA-RBF 模型的性能,采用网络流量文库: <http://newsfeed.ntcu.net/~news/2010/> 主节点路由器 2010 年 1 月 1 日到 2 月 20 日每小时网络流量数据进行仿真实验,共收集到 1000 个数据.为了防止预测结果偶然性,采用前 800 个数据作为训练集建立网络流量预测模型,最后 200 个数据作

为测试集,对建立的模型性能验证. 由于 RBF 神经网络对[0,1]数据最为灵敏,实际网络流量是具有随机性,变化幅度大,为了提高 RBF 神经网络学习速度,在训练前对网络流量数据进行归一化处理.

采用 VC++ 语言实现 GA-BRF 算法,采用 CC-RBF 神经网络模型进行对比实验. 模型性能的评价指标为均方误差 MSE 和平均绝对百分误差 (mean absolute percentage error, MAPE)<sup>[8]</sup>.

4.2 结果与分析

(1) 模型的拟合结果对比

将归一化后的训练样本分别输入到采用 GA-RBF 和 CC-RBF 模型中进行训练,采用 GA 对模型的参数进行优化,得到的参数分别如表 1 所示. 然后采用最优参数建立相应的网络流量 GA-RBF 和 CC-RBF 预测模型,然后采用建立的模型对网络流量的训练集进行拟合,拟合结果如图 2 所示.

表 1 模型的参数优化结果

| 参数       | CC-RBF | GA-RBF |
|----------|--------|--------|
| $\tau$   | 1      | 1      |
| $m$      | 4      | 7      |
| $w$      | 2.0    | 2.0    |
| $c$      | 10.5   | 6.4    |
| $\sigma$ | 1.0    | 2.0    |

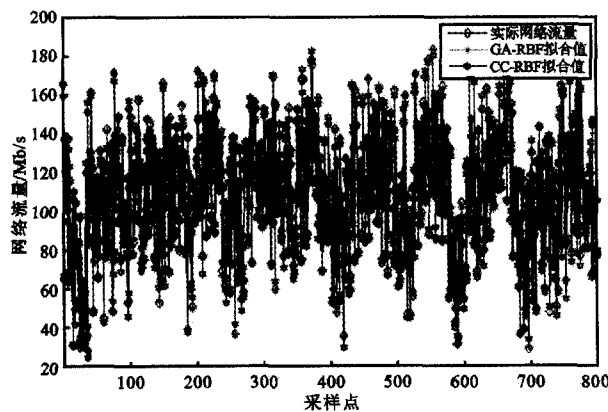


图 2 两种模型训练集的拟合结果比较

从图 2 可知,相对于参数单独优化的 CC-RBF, GA-RBF 的网络流量预测值更接近实际的流量变化数值,拟合的精度更高,对比结果初步表明,基于参数同步优化的 GA-RBF 模型对网络流量进行预测可行的,能够提高网络流量的拟合精度.

(2) 模型的预测能力对比

评价一个预测模型性能的优劣主要考察其预测能力而非拟合能力,因此采用 CC-RBF 和 GA-RBF 对训练集建立的最优预测模型对测试集进行预测,分进行一步预测和多步预测方法,其结果分别如图 3、图 4 所示. 对比结果可知,相对于 CC-RBF 模型, GA-RBF 预测精度均有所提高. 尤其对于多步预测, GA-RBF 预测误差较低,而 CC-RBF 预测误差比较,说明 GA-RBF 的预测性能比较稳定,具有更强的鲁棒性,较好的刻画了现代网络流量时变性、非线性、空变性等变化特点,是一种高精度、预测结果可靠的网络流量预测模型.

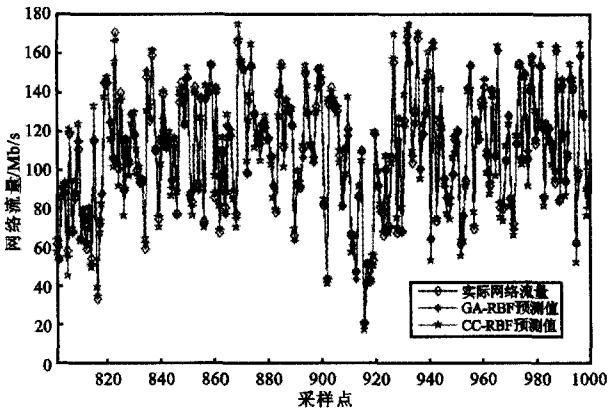


图 3 两种模型测试集的一步预测结果对比

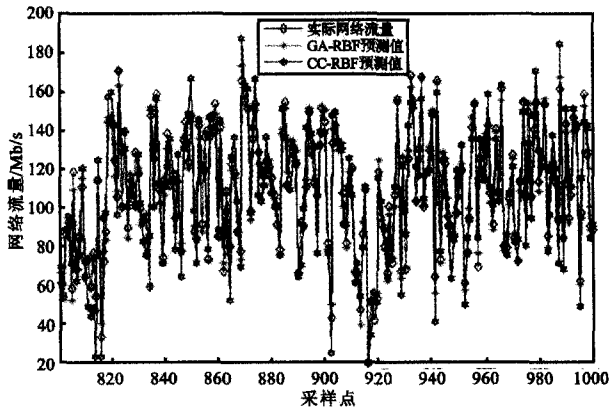


图 4 两种模型测试集的多步预测结果对比

(3) 两种模型的评价指标对比

两种模型的性能评价指标如表 2 所示. 从表 2 可知:①对于训练集和测试集,GA-RBF 拟合和预测误差都有所降低,这表明,对相空间重构和 RBF 神经网络参数同步优化的思想是正确的,能够提高网络流量预测模型的各项性能,更好的刻画现代网络流量的变化趋势,GA-RBF 模型更适合于长期预测. ②从运行时间可知,GA-RBF 的运行速度大大提高,可以很好的满足网络流量预测的实时性需求.

表 2 两种网络流量预测模型的整体性能对比

|        | GA-RBF |            |           | CC-RBF |            |           |
|--------|--------|------------|-----------|--------|------------|-----------|
|        | 训练集    | 预试集(1 步预测) | 预试集(多步预测) | 训练集    | 预试集(1 步预测) | 预试集(多步预测) |
| MSE    | 0.65   | 61.68      | 17.2      | 17.5   | 3.98       | 170.5     |
| MAPE   | 0.69%  | 1.71%      | 11.03%    | 3.27%  | 6.43%      | 3.44%     |
| 运行时间/s | 30     | 2.5        | 4.5       | 65     | 3.3        | 5.7       |

5 结束语

文中提出了一种 GA—RBF 的网络流量预测模型. 与其它模型相比,预测结果可以很好的满足网络流量预测高精度、实时性要求,泛化能力更强.

参考文献:

[1] Nguyen T T, Armitage G. A survey of techniques for internet traffic classification using machine learning [J]. IEEE Communications Surveys and Tutorials, 2008, 10(4):56-76.

[2] 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): 352-356.

[3] 冯海亮,陈涤,林青家,等. 一种基于神经网络的网络流量组合预测模型[J]. 计算机应用,2006, 26(9): 2206-22208.

[4] 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-

1663.

[5] 王鹏,刘渊. 基于改进的 QPSO 训练 BP 网络的网络流量预测[J]. 计算机应用研究,2009,26(1):299-301.

[6] 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):155-15

[7] 邹柏贤,姚志强. 一种网络流量平稳化方法[J]. 通信学报,2004,25(8):14-23.

[8] 张颖璐. 基于遗传算法优化支持向量机的网络流量预测[J]. 计算机科学,2008,35(5):177-180.

[9] 任伟,蒋兴浩,孙铎锋. 基于 RBF 神经网络的网络安全态势预测方法[J]. 计算机工程与应用,2006(31):136-138.

作者简介:

赵清艳 女,(1972—),硕士,工程师,系统分析师. 研究方向为计算机软件、虚拟仪器.

(上接第 131 页)

[5] 国家标准. GB—T 11318. 7—1996 电视和声音信号的电缆分配系统设备与部件第 7 部分放大器通用规范 [S]. 1996.

[6] 国家标准. GB—T 11318. 8—1996 电视和声音信号的电缆分配系统设备与部件第 8 部分干线放大器通用规范[S]. 1996.

作者简介:



关雪梅 女,(1976—),硕士,讲师. 研究方向为计算机网络应用与数据通信.