

# 基于局域相关向量机回归模型的小尺度网络流量的非线性预测\*

孟庆芳<sup>†</sup> 陈月辉 冯志全 王枫林 陈珊珊

(济南大学信息科学与工程学院, 济南 250022)

(山东省网络环境智能计算技术重点实验室, 济南 250022)

(2013年1月10日收到; 2013年4月24日收到修改稿)

基于非线性时间序列局域预测法与相关向量机回归模型, 本文提出了局域相关向量机预测方法, 并应用于预测实际的小尺度网络流量序列. 应用基于信息准则的局域预测法邻近点的选取方法来选取局域相关向量机回归模型的邻近点个数. 对比分析了局域相关向量机预测法、前馈神经网络模型与局域线性预测法对网络流量序列的预测性能, 其中前馈神经网络模型的参数采用粒子群优化算法来优化. 实验结果表明: 邻近点优化后的局域相关向量机回归模型能够有效地预测小尺度网络流量序列, 归一化均方误差很小; 局域相关向量机回归模型生成的时间序列具有与原网络流量时间序列相一致的概率分布; 局域相关向量机回归模型的预测精度好于前馈神经网络模型的与局域线性预测法的.

**关键词:** 小尺度网络流量, 非线性时间序列预测方法, 局域预测法, 相关向量机回归模型

**PACS:** 05.45.Tp, 05.45.Pq, 05.45.Ac

**DOI:** 10.7498/aps.62.150509

## 1 引言

网络流量的特性刻画与模型建立一直是网络技术研究中的一个热点问题. 一个具有精确刻画能力的模型的建立, 对分析、理解和仿真网络的动态行为, 对指导网络流量均衡控制方案的设计, 对相关网络路由、交换和管理设备及其支撑软件的开发都具有基础性的重要意义.

复杂性是网络信息传输的一个关键问题. 网络流量的复杂性表现为多种形式, 例如长相关特性和自相似性. 现有研究表明不同尺度的网络流量表现出不同的性质. 网络流量的复杂性表明网络流量不能仅在现有的网络流量建模预测的框架内分析<sup>[1-3]</sup>. 应该把适合网络流量特性的新的信号处理理论与方法应用于网络流量的特性刻画、建模预测和控制等<sup>[4,5]</sup>.

近年来, 通信与网络技术的飞速发展使网络流量的特性发生重大变化, 网络流量特性分析与模型建立已从大尺度上转移到小尺度上来. 已有研究表明网络流量在小尺度上的特性不同于其在大尺度上的特性<sup>[6,7]</sup>. 因此, 原有的大尺度流量模型不再适用于小尺度流量模型.

基于非线性动力学理论的非线性时间序列分析提供了处理非线性不规则时间序列新的方法和思路<sup>[8-23]</sup>. 已有研究表明网络流量序列中存在非线性确定性成份<sup>[24,25]</sup>. 本文将应用非线性时间序列预测方法来预测小尺度网络流量序列. 本文的研究将对小尺度网络流量有更加深入的认识.

非线性时间序列预测方法可被分为两类: 全局预测法<sup>[9-14]</sup>与局域预测法<sup>[15-23]</sup>. 近年来, 神经网络<sup>[9-12]</sup>、支持向量机(SVM)<sup>[13,14]</sup>等全局模型已广泛的应用于非线性时间序列预测中. 前馈神经网络是应用最广泛的一种神经网络, 文献<sup>[26]</sup>研究

\* 国家自然科学基金(批准号: 61201428, 61070130, 61173079)、山东省自然科学基金(批准号: ZR2010FQ020, ZR2011FZ003)、山东省优秀中青年科学家科研奖励基金(批准号: BS2009SW003)和中国博士后科学基金(批准号: 20100470081)资助的课题.

<sup>†</sup> 通讯作者. E-mail: ise\_mengqf@ujn.edu.cn

了基于前馈神经网络模型的网络流量预测问题. 柔性神经树作为一种特殊的多层前馈神经网络, 大大改变了传统的神经网络的结构设计和参数优化问题, 其结构和参数由优化算法自动的设计和优化得到 [27,28]. 文献 [29] 把柔性神经树模型应用于对网络流量序列的预测.

相关向量机 (relevance vector machine, RVM) 是由 Tipping 于 2000 年提出的基于总体贝叶斯框架下的稀疏概率模型, 是近年来统计学习领域的重要的研究热点之一 [30]. 相关向量机拥有支持向量机 (SVM) 的工作性能, 同时具有支持向量机所不具备的优点, 例如 RVM 具有优于 SVM 的稀疏性; RVM 的核函数不必满足 Mercer 条件; RVM 在权值上引进了超参数, 从而大大降低了计算的复杂度等 [30]. 目前, RVM 模型已在建模预测与分类方面有所应用 [31,32].

局域预测法是非线性时间序列预测最常用的方法之一. 基于非线性时间序列局域预测法与相关向量机回归模型, 本文提出了局域相关向量机预测方法, 并应用基于信息准则的局域预测法邻近点的选取方法 [23] 来选取局域相关向量机回归模型的邻近点个数, 最后把邻近点优化后的局域相关向量机回归模型应用于对实际的小尺度网路流量序列的建模预测. 对比分析了局域相关向量机预测法、前馈神经网络模型与局域线性预测法对网络流量序列的预测性能, 其中前馈神经网络模型的参数采用粒子群优化算法来优化.

## 2 小尺度网络流量序列的非线性预测

### 2.1 相关向量机回归模型

给定数据集  $\{x_n, t_n\}_{n=1}^N$ , 其中  $x_n$  为输入向量,  $t_n$  为标量输出, 假定目标输出是对下列含加性噪声的模型采样得到的:

$$t_n = y(x_n; w) + \xi_n = \sum_{i=1}^N w_i \varphi_i(x) + w_0 + \xi_n, \quad (1)$$

其中,  $w$  为权参数向量,  $w = (w_0, w_1, \dots, w_N)$ ;  $\xi_n$  为独立同分布的零均值高斯白噪声, 其方差为  $\sigma^2$ ;  $\varphi_i(x)$  为非线性基函数,  $\varphi_i(x) \equiv K(x, x_i)$ ,  $K(\bullet)$  为核函数. 因此  $p(t_n | x) = N(t_n | y(x_n), \sigma^2)$ , 表示  $t_n$  的高斯分布, 其中均值为  $y(x_n)$ 、方差为  $\sigma^2$ . 由于假设  $t_n$  相互

独立, 全部数据集的似然估计可写为

$$p(t | w, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2} \|t - \Phi w\|^2\right\}, \quad (2)$$

式中,  $t = (t_1, t_2, \dots, t_N)^T$ ;  $\Phi$  为  $N \times (N+1)$  矩阵,  $\Phi = [\varphi(x_1), \varphi(x_2), \dots, \varphi(x_N)]^T$ , 其中  $\varphi(x_n) = [1, K(x_n, x_1), K(x_n, x_2), \dots, K(x_n, x_N)]^T$ .

由于模型中存在较多参数, 采用最大似然估计由 (2) 式得到的  $w$  和  $\sigma^2$  易于导致过拟合, 因此, 采用稀疏贝叶斯原理对  $w$  赋予零均值高斯先验分布得

$$p(w | \alpha) = \prod_{i=0}^N N(w_i | 0, \alpha_i^{-1}), \quad (3)$$

其中  $\alpha$  为  $N+1$  维的超参数向量. 这样, 每一个权重就单独地对应一个超参数, 从而控制先验分布对各参数的影响, 以确保相关向量模型的稀疏性.

在以上定义了先验概率分布与似然分布之后, 根据贝叶斯原理, 可以求得所有未知参数的后验概率分布为

$$p(w | t, \alpha, \sigma^2) = (2\pi)^{-(N+1)/2} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(w - \mu)^T \Sigma^{-1}(w - \mu)\right\}, \quad (4)$$

式中, 后验协方差矩阵与均值分别为

$$\Sigma = (\sigma^{-2} \Phi^T \Phi + A)^{-1}, \quad (5)$$

$$\mu = \sigma^{-2} \Sigma \Phi^T t, \quad (6)$$

其中  $A = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$ .

基于最大期望超参数估计,  $\alpha$  和  $\sigma^2$  的值可以通过迭代算法求得, 即

$$\alpha_i^{\text{new}} = \frac{\gamma_i}{\mu_i^2}, \quad (7)$$

式中,  $\mu_i$  为第  $i$  个后验平均权值,  $\gamma_i$  定义如下式所示:

$$\gamma_i \equiv 1 - \alpha_i N_{ii},$$

式中  $N_{ii}$  为由 (5) 式得到的后验权值协方差矩阵的第  $i$  个对角元素, 其中  $\alpha$  和  $\sigma^2$  取当前的值.

噪声方差  $\sigma^2$  通过迭代算法求得, 即

$$(\sigma^2)^{\text{new}} = \frac{\|t - \Phi \mu\|^2}{N - \sum_i \gamma_i}, \quad (8)$$

注意上式分母中  $N$  指样本数据的个数, 而不是基函数的个数.

在超参数估计的收敛过程中, 基于权值的后验概率分布进行预测, 其限制条件为  $\alpha_{\text{MP}}$  和  $\sigma_{\text{MP}}^2$  取

最大值. 若给定新的输入值  $x_*$ , 对应的预测目标值为  $t_*$ , 则相应预测输出的概率分布服从高斯分布

$$p(t_*|t, \alpha_{MP}, \sigma_{MP}^2) = N(t_*|y_*, \sigma_*^2), \quad (9)$$

其中

$$y_* = \mu^T \Phi(x_*), \quad (10)$$

$$\sigma_*^2 = \sigma_{MP}^2 + \Phi(x_*)^T \Sigma \Phi(x_*). \quad (11)$$

那么  $y_*$  可作为  $t_*$  的预测值.

## 2.2 非线性时间序列局域预测方法

从观测时间序列预测系统的时间演化序列是时间序列分析的最重要的经典问题之一. 随着混沌理论及其应用技术研究的发展, 提出了大量的非线性时间序列预测方法 [9-22] 来预测不规则时间序列, 它们可被分为两类, 全局预测法和局域预测法. 全局预测法 [9-14] 利用全部的去信息来预测未来值, 用全部已知数据来拟合动力方程. 例如用神经网络、支持向量机或相关向量机等非线性模型建立的全局预测模型, 它通过全部的输入-输出对非线性模型进行训练. 但当新的数据加入已知样本时, 全局预测模型须重新估计其参数, 因此全局预测法计算量较大. 并且因为吸引子的结构非常复杂, 所以拟合全局动力方程的难度也往往较大. 而局域预测法 [15-22] 仅利用部分的过去信息来预测未来值, 局域动力方程较容易拟合, 且计算量较小. Farmer 和 Sidorowich 早已证明, 在相同的嵌入维数下, 局域预测法的效果比全局预测法更好 [15]. 正是因为如此, 局域预测法受到了学术界的重视和关注, 成为非线性时间序列预测最常用的方法之一.

局域预测的首要步骤是在重构的相空间中选取当前延迟矢量  $X(n)$  的邻近点. 根据重构轨迹, 计算当前延迟矢量  $X(n)$  与前面的  $n-1$  个延迟矢量  $X(i), i = 1, 2, 3, \dots, n-1$  的距离

$$d(i) = \|X(i) - X(n)\|_2, \quad (12)$$

可得到  $k$  个最近的邻近点  $X(n_i)(i = 1, 2, \dots, k)$ . 局域线性预测模型为

$$\hat{x}_{n+T} = a_n X(n) + b_n. \quad (13)$$

其中  $a_n$  为  $1 \times m$  维行向量,  $b_n$  为常数,  $T$  为预测步长,  $m$  为嵌入维数. 局域线性预测法利用当前延迟矢量  $X(n)$  的邻近点  $X(n_i)(i = 1, 2, \dots, k)$  来估计参

数  $a_n$  和  $b_n$  的值. 参数  $a_n$  和  $b_n$  的值可通过最小化下式来得到:

$$\min \sum_{i=1}^k |x_{n_i+T} - a_n X(n_i) - b_n|^2. \quad (14)$$

局域线性预测法用线性映射来近似当前延迟矢量  $X(n)$  与未来值  $\hat{x}_{n+T}$  的演化方程, 并用  $X(n)$  的邻近点  $X(n_i)(i = 1, 2, \dots, k)$  来近似该线性映射.

## 2.3 小尺度网络流量序列的局域预测

时至今日, 国内外对局域预测法的研究仍然在继续进行. 但是, 许多研究者提出局域非线性模型与局域线性模型有近似的预测性能, 因此许多研究者在局域预测中仅使用线性模型 [15]. 近年来, 许多研究者尝试在局域预测法中用非线性模型来代替线性模型 [20-22].

邻近点个数是局域预测法的非常重要的参数, 它决定局域模型的预测精度和计算量. 本文笔者认为在局域线性假设条件的限制下局域非线性模型具有与局域线性模型相类似的预测性能. 本文提出为了在局域预测法中发挥局域非线性模型非线性拟合的优势, 为了提高局域非线性模型的预测精度, 局域非线性模型的邻近点个数的值可以大于局域线性模型的. 同时, 局域非线性模型不受局域线性这一假设条件的限制, 因此其邻近点个数完全可以比局域线性模型更大. 但是, 另一方面, 如果邻近点的个数  $k$  取得太大, 不仅局域非线性模型的计算量将大大增加, 计算时间大大增加, 而且有可能出现“过拟合”(overfitting) 的现象, 反而降低预测的精度. 所以, 对于局域非线性预测模型, 在满足预测精度较高的条件下, 邻近点也不易过多.

本文用相关向量机回归模型代替局域的线性模型, 并用邻近点训练该局域相关向量机回归模型, 提出局域相关向量机预测法. 即把  $X(n_i)(i = 1, 2, \dots, k)$  作为相关向量机预测模型中的输入值,  $x_{n_i+T}(i = 1, 2, \dots, k)$  作为输出值训练相关向量机, 得到对应的相关向量及权重系数(即动力方程一个子集的逼近); 然后利用训练得到的局域相关向量机, 计算得到对未来值的预测值  $\hat{x}_{n+T}$ .

本文应用局域相关向量机回归模型来预测实测网络流量数据, 并应用基于 BIC 准则的局域预测邻近点的选取方法 [23] 来选取局域相关向量机预测模型的邻近点个数, 即最小化下式:

$$N(k) = \ln \varepsilon(k)^2 + (k+1) \frac{\ln L}{L}, \quad (15)$$

其中  $\epsilon(k)^2$  是预测误差的归一化的方差,  $L$  为预测序列的长度. 最后把邻近点优化后的局域相关向量机回归模型应用于对实际的小尺度网路流量序列的建模预测.

### 3 实验结果及分析

本文的网络流量数据取自 Lawrence Berkeley Laboratory 发布的 TCP 流量数据, 这些数据包括了该实验室与世界其他地方通过广域网进行联系的数据, 本文使用数据包 DEC-Pkt1, 该流量为 1 h 的广域网 TCP 流量, 通信双方为 Digital Equipment Corporation 与世界其他地方, 时间精度为 1 ms<sup>[33]</sup>. 本文将应用非线性时间序列预测方法来预测聚类时间为 0.1 s 的 DEC-Pkt1 小尺度网络流量数据, 既 0.1 s 时间间隔内包到达的数量, 如图 1 所示.

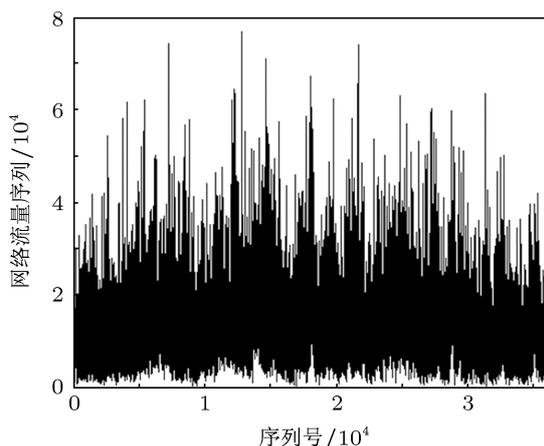


图 1 聚类时间为 0.1 s 的小尺度网络流量序列

实测网路流量数据通常不可避免的被随机噪声所污染. 众所周知小波软阈值去噪方法具有降低分析信号的随机噪声成份的优良性能. 为了降低网络流量序列的随机噪声部分, 本文应用小波软阈值去噪方法来滤除网络流量数据的随机噪声成分. 把降低噪声成份后的网路流量序列作为分析信号, 并把分析序列归一化到区间 [0,1].

应用基于 BIC 准则的局域预测邻近点的选取方法<sup>[19]</sup>来选取局域相关向量机预测模型的邻近点个数, 实验结果如图 2 所示. 从图 2 可见, 当  $k = 47$  时,  $N(k)$  取到最小值, 因此局域相关向量机预测模型的邻近点个数取为 47.

应用邻近点优化后的局域相关向量机预测方法来预测网络流量序列, 其中邻近点个数取为 47.

该组小尺度网络流量序列的长度为 36000, 把前 33000 个数据作为训练样本, 后 3000 个数据作为测试样本. 用归一化均方误差 (NMSE) 来评测预测精度, NMSE 定义为

$$NMSE = \frac{(1/M) \sum_{i=1}^M |x(N+i) - x_p(N+i)|^2}{(1/N) \sum_{i=1}^N [x(i) - \bar{x}]^2}, \quad (16)$$

其中  $M = 3000$ ,  $N = 33000$ . 相关向量回归模型的核函数取为高斯核函数, 其中高斯核函数的宽度值取为 0.015, 迭代次数取为 100. 局域相关向量机预测方法对网络流量序列的预测结果如图 3、图 4 所示, 其中图 3 为网络流量的真实值与预测值, 图 4 为预测误差.

从图 3 和图 4 可见, 邻近点优化后的局域相关向量机预测方法能够有效地预测网络流量序列, 预测精度高, 归一化均方误差很小, 仅为 0.0085.

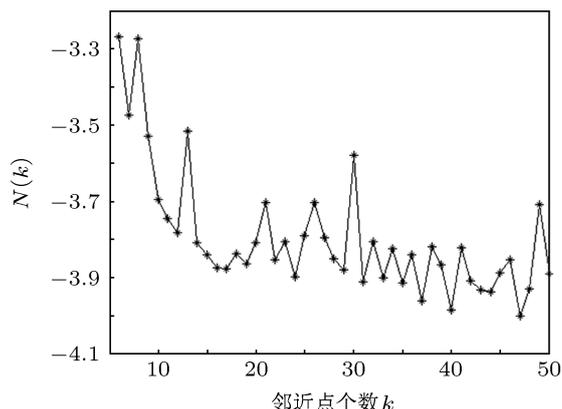


图 2 对网路流量序列, 局域相关向量机预测方法的基于 BIC 准则的邻近点个数  $(N(k) - k)$

为了对比分析, 本文分别采用前馈神经网络模型与局域线性预测法对该组网络流量序列进行预测. 其中前馈神经网络全局模型的输入层个数取为 10, 隐含层个数取为 12, 输出层个数取为 1. 采用粒子群优化算法 (PSO) 来优化该前馈神经网络模型的参数. 局域线性预测法的嵌入维数  $m$  取为 5, 邻近点个数取为  $2m + 1$  个, 即 11 个. 该前馈神经网络全局模型与局域线性预测法对该组序列后 3000 个数据点的预测误差的归一化均方误差分别为  $NMSE = 0.0732$ ,  $NMSE = 0.0143$ , 如表 1 所示. 由此可见, 对该组实际的网路流量序列, 本文提出的局域相关向量机预测法的预测性能好于前馈神经网络模型的与局域线性预测法的.

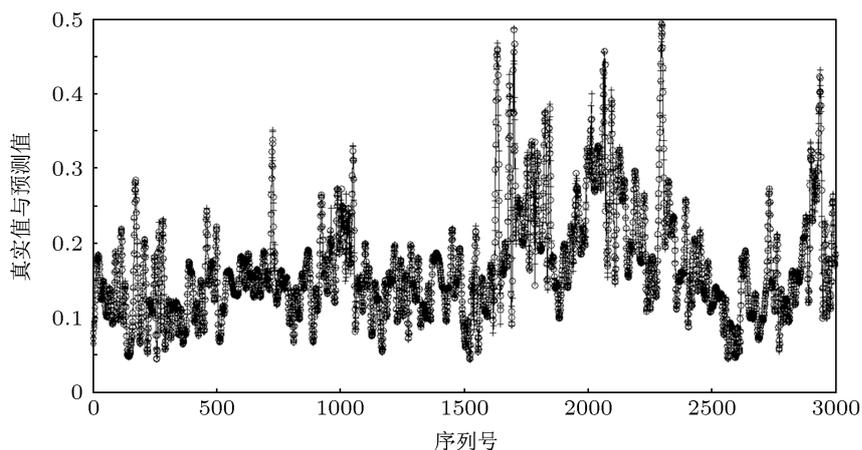


图3 局域相关向量机预测方法对网络流量序列的真实值与预测值,  $k = 47$  (圆圈为真实值, 加号为预测值)

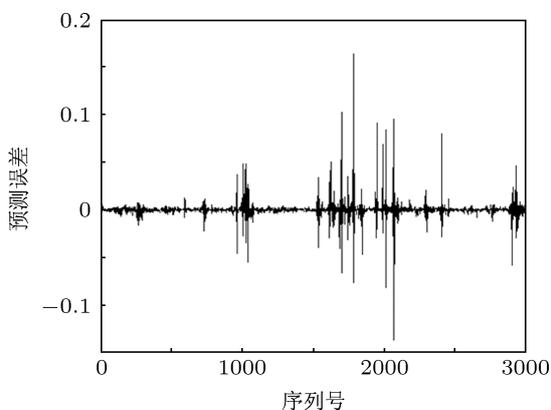


图4 局域相关向量机预测方法对网络流量序列的一步预测误差,  $NMSE = 0.0085$

表1 对网路流量序列预测的归一化均方差 NMSE

方法	NMSE
前馈神经网络模型 +PSO	0.0732
局域线性预测法	0.0143
局域相关向量机预测法	0.0085

对比分析原网络流量序列和局域相关向量回归模型生成时间序列的统计分布特性, 实验结果如图5所示; 图中横坐标为网络流量的真实值和预测值, 纵坐标为真实值或预测值的概率分布. 局域相关向量机预测法的预测误差的统计分布特性如图6所示, 图中横坐标为预测误差值, 纵坐标为预测误差值的概率分布.

从图5可见, 局域相关向量机回归模型生成的时间序列具有与原网络流量时间序列非常近似的概率分布.

从图6可见, 局域相关向量机回归模型对小尺度网络流量的预测误差主要集中在零值附近, 且绝

对值较大的预测误差的概率非常小.

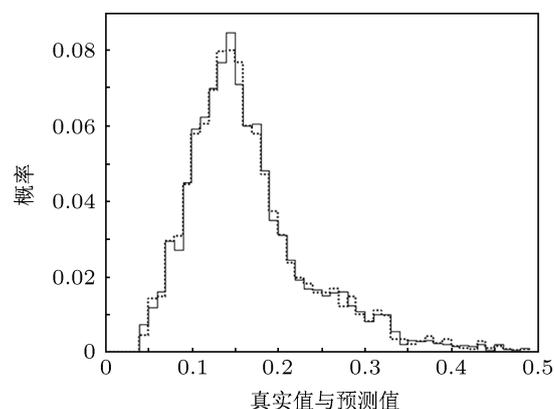


图5 网络流量真实值和预测值的概率分布 (实线为真实值的概率分布, 点线为预测值的概率分布)

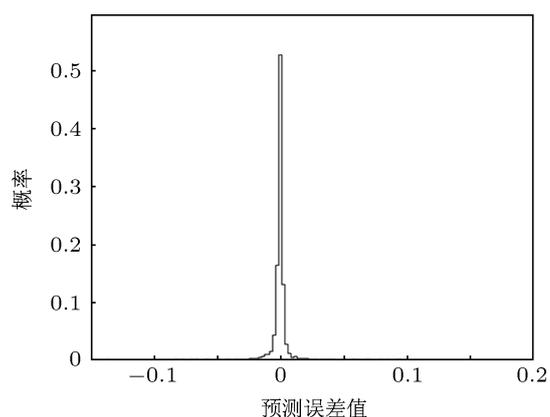


图6 局域相关向量机预测法的预测误差的概率分布

## 4 结论

本文应用非线性时间序列局域预测方法来预测实测的网络流量序列. 本文的研究表明非线性时间序列预测方法的成功应用将使研究者们对网络

流量数据的主要特征有更加深入的认识. 为了降低网络流量序列中的随机噪声成分, 应用小波软阈值去噪方法来滤除网络流量数据的随机噪声成分. 应用局域相关向量机预测方法来预测网络流量序列. 并应用基于 BIC 准则的局域预测法邻近点的选取方法来选取局域相关向量机回归模型的邻近点个数. 实验结果表明邻近点优化后的局域相关向量机

预测方法能够有效地预测网络流量序列, 预测精度高, 归一化均方误差很小; 局域相关向量回归模型生成的时间序列具有与原网络流量时间序列相一致的分布; 局域相关向量机预测法的预测精度好于粒子群优化算法优化的前馈神经网络模型与局域线性预测法.

- [1] Orosz G, Krauskopf B, Wilson R E 2005 *Physica D* **211** 277
- [2] Orosz G, Wilson R E, Krauskopf B 2004 *Phys. Rev. E* **70** 026207
- [3] Gasser I, Siritto G, Werner B 2004 *Physica D* **197** 222
- [4] Doulamis A D, Doulamis N D, Kollias S D 2003 *IEEE Trans. on Neur. New.* **14** 150
- [5] Xie Y B, Wang W X, Wang B H 2007 *Phys. Rev. E* **75** 026111
- [6] Zhang Z L, Ribeiro V J, Moon S, Diot C 2003 *IEEE INFOCOM* **3** 1826
- [7] Uglic S 2004 *ACM SIGCOMM computer communications review* **34**
- [8] Kantz H, Schreiber T 2003 *Nonlinear Time Series Analysis* (Second Edition) (Cambridge: Cambridge University Press)
- [9] Small M, Tse C K 2002 *Phys. Rev. E* **66** 066701
- [10] Karunasinghe D S K, Liong S Y 2006 *Journal of Hydrology* **323** 92
- [11] Ma Q L, Zheng Q L, Peng H, Zhong T W, Qin J W 2008 *Chin. Phys. B* **17** 536
- [12] Ma Q L, Zheng Q L, Peng H, Qin J W 2009 *Acta Phys. Sin.* **58** 1410 (in Chinese) [马千里, 郑启伦, 彭宏, 覃姜维 2009 物理学报 **58** 1410]
- [13] Liu H, Liu D, Deng L F 2006 *Chin. Phys.* **15** 1196
- [14] Chen Q, Ren X M 2010 *Acta Phys. Sin.* **59** 2310 (in Chinese) [陈强, 任雪梅 2010 物理学报 **59** 2310]
- [15] Farmer J D, Sidorowich J J 1987 *Phys. Rev. Lett.* **59** 845
- [16] Giona M, Lentini F, Cimagalli V 1991 *Phys. Rev. A* **44** 3496
- [17] Kugiumtzis D, Lingjarde O C, Christophersen N 1998 *Physica D* **112** 344
- [18] Ragwitz M, Kantz H 2002 *Phys. Rev. E* **65** 056201
- [19] Meng Q F, Peng Y H 2007 *Phys. Lett. A* **370** 465
- [20] Li H C, Zhang J S 2005 *Chin. Phys. Lett.* **22** 2776
- [21] Zhang J S, Dang J L, Li H C 2007 *Acta Phys. Sin.* **56** 67 (in Chinese) [张家树, 党建亮, 李恒超 2007 物理学报 **56** 67]
- [22] Du J, Cao Y J, Liu Z J, Xu L J, Jiang Q Y, Guo C X, Lu J G 2009 *Acta Phys. Sin.* **58** 5997 (in Chinese) [杜杰, 曹一家, 刘志坚, 徐立中, 江全元, 郭创新, 陆金桂 2009 物理学报 **58** 5997]
- [23] Meng Q F, Peng Y H, Qu H J, Han M 2008 *Acta Phys. Sin.* **57** 1423 (in Chinese) [孟庆芳, 彭玉华, 曲怀敬, 韩民 2008 物理学报 **57** 1423]
- [24] Shang P, Li X, Kamae S 2005 *Chaos Solitons Fractals* **25** 121
- [25] Shang P, Li X, Kamae S 2006 *Phys. Lett. A* **357** 314
- [26] Akritas P, Akishin P G, Antoniou I, Bonushkina A Y, Drossinos I 2002 *Chaos Solit. and Frac.* **14** 595
- [27] Chen Y H, Yang B, Dong J W, Abraham A 2005 *Information Sciences* **174** 219
- [28] Chen Y H, Yang B, Abraham A 2007 *Neurocomputing* **70** 697
- [29] Chen Y H, Yang B, Meng Q F 2012 *Applied Soft Computing* **12** 274
- [30] Tipping M E 2001 *Journal of Machine Learning Research* **3** 211
- [31] Zio E, Maio F D 2012 *Expert Systems with Applications* **39** 10681
- [32] Han M, Zhao Y, Yang X L, Lin D 2011 *Control Theory and Applications* **28** 343 (in Chinese) [韩敏, 赵耀, 杨溪林, 林东 2011 控制理论与应用 **28** 343]
- [33] Internet traffic archive. <http://ita.ee.lbl.gov/>

# Nonlinear prediction of small scale network traffic based on local relevance vector machine regression model\*

Meng Qing-Fang<sup>†</sup> Chen Yue-Hui Feng Zhi-Quan  
Wang Feng-Lin Chen Shan-Shan

(School of Information Science and Engineering, University of Jinan, Jinan 250022, China)

(Shandong Provincial Key laboratory of Network Based Intelligent Computing, Jinan 250022, China)

(Received 10 January 2013; revised manuscript received 24 April 2013)

## Abstract

Based on the nonlinear time series local prediction method and the relevance vector machine regression model, the local relevance vector machine prediction method is proposed and applied to predict the small scale traffic measurement data, and the BIC-based neighbor point selection method is used to choose the number of nearest-neighbor points for the local relevance vector machine regression model. We also compare the performance of the local relevance vector machine regression model with the feed-forward neural network optimized by particle swarm optimization for the same problem. Experimental results show that the local relevance vector machine prediction method whose neighboring points have been optimized can effectively predict the small scale traffic measurement data, can reproduce the statistical features of real small scale traffic measurements, and the prediction accuracy of the local relevance vector machine regression model is superior to that of the feedforward neural network optimized by PSO and the local linear prediction method.

**Keywords:** small scale network traffic, nonlinear time series prediction method, local prediction method, relevance vector machine regression model

**PACS:** 05.45.Tp, 05.45.Pq, 05.45.Ac

**DOI:** 10.7498/aps.62.150509

---

\* Project supported by the National Natural Science Foundation of China (Grant Nos. 61201428, 61070130, 61173079), the Natural Science Foundation of Shandong Province, China (Grant Nos. ZR2010FQ020, ZR2011FZ003), the Shandong Distinguished Middle-aged and Young Scientist Encourage and Reward Foundation, China (Grant No. BS2009SW003), and the China Postdoctoral Science Foundation (Grant No. 20100470081).

<sup>†</sup> Corresponding author. E-mail: ise\_mengqf@ujn.edu.cn