

基于最大相关熵准则的网络流量预测^①

曲 桦^{②*} 马文涛^{③*} 赵季红* 王 涛**

(* 西安交通大学电信学院 西安 710049)

(** 西安邮电大学通信与信息工程学院 西安 710121)

摘要 为提高网络流量预测的精度,针对网络流量的非线性特征提出了一种基于新的误差评价准则——最大相关熵准则(MCC)的网络流量预测方法。该方法使用 MCC 对 Elman 神经网络进行训练。该评价准则是基于新的相似度函数——广义相关熵(correntropy)函数的概念建立的,此相似度函数以误差概率密度函数的 Parzen 窗估计和瑞利熵为基础。同时结合 MCC 和最小均方误差(MMSE)准则提出了一种混合的评价准则 MCC-MMSE。针对网络流量的非线性、非高斯性、突发性等特性,分别以 MCC、MCC-MMSE 准则进行了 Elman 神经网络的训练,使用训练好的神经网络进行网络流量预测,仿真结果表明预测结果的精度优于以 MMSE 为准则的 Elman 神经网络的预测结果。

关键词 最大相关熵准则(MCC), 最小均方误差(MMSE), Elman 神经网络, 网络流量预测

0 引言

网络流量是衡量网络负荷和网络当前状态的重要参数,进行网络流量预测对网络管理有重要意义。通过对网络流量建模预测出在未来超过流量阈值的可能性和时间,提早做出防范,可大大提高网络管理性能和有效改善 QoS,因而基于网络流量预测的资源分配机制已得到重视^[1-3]。目前已有许多网络流量预测算法,主要的有基于时间序列模型的方法^[4]和基于神经网络的方法^[5-8]。基于时间序列模型的方法虽然预测精度较高,但只能处理平稳过程,且因为是线性预测模型,对突发性不敏感,而实环境中的网络流量呈现出相当明显的非线性、长短相关性并存、突发性等特性^[9],使得传统预测模型的误差较大。神经网络对非线性函数关系具有良好的逼近能力,适合描述网络流量的非线性、长相关性等。已有神经网络预测模型在网络学习过程中使用的评价函数都基于最小均方误差(MMSE)准则,MMSE 准则考虑了预测误差分布的二阶距,对有高斯特性的数据而言其有效性非常高。然而,在大多数情况下,网

络流量预测的误差分布往往是非高斯的,因此基于 MMSE 的神经网络用于网络流量预测就会存在一定的缺陷。为了很好地处理具有非高斯特性的误差分布问题,美国佛罗里达大学神经网络实验室结合信息论中熵的概念和相关函数提出了一种新的相似度评价函数——广义相关熵^[10,11](generalized correlation entropy function, correntropy)函数,而且以熵和 Parzen 窗估计为基础,提出了一种新的误差评价准则——最大相关熵准则(maximum correntropy criterion, MCC),它以最大化误差分布的信息熵进行网络的训练。此评价准则考虑到了数据的高阶统计量,对于非高斯和非线性特征数据的处理具有一定的鲁棒性,因而很适合于网络流量的预测。本文基于以上分析,首次将 MCC 引入到 Elman 神经网络,提出了基于 MCC 和 MCC 与 MMSE 准则混合的两种用于预测网络流量 Elman 神经网络预测模型。

1 学习指标评价准则

对于自适应系统的学习过程需要代价准则来评价学习的效果,目前经典的准则为 MMSE 准则。这

① 国家自然科学基金(61071126)和国家无线重大专项(2010ZX03004-001, 2010ZX03004-002, 2011ZX03002-001)资助项目。

② 男,1961 年生,博士,教授;研究方向:通信网络生存性;E-mail: qh@mail.xjtu.edu.cn

③ 通讯作者, E-mail: mwt7980@yahoo.com.cn

(收稿日期:2012-02-23)

里对本文用到的三种评价准则进行描述。

1.1 最大互相关熵准则(MCC)

Principe 教授及其团队从信息论和随机过程的相关性方面进行了深入研究,提出了广义相关熵(correntropy)的概念^[10],其可作为一个新的相似度函数。对于任意两个随机变量 Z 和 Y,其定义如下:

$$V(Z, Y) = E(k(Z - Y))$$

$$= \iint_{x,y} k(z - y) f_{ZY}(z, y) dz dy$$

$$= \int_e k(e) f_E(e) de \quad (1)$$

其中误差 $e = z - y$, $k(\cdot)$ 表示核函数,常用高斯核函数,即 $k(\cdot) = G_\sigma(\cdot) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\cdot)^2}{2\sigma^2}}$, $\sigma > 0$, 由此得到 correntropy 的新形式:

$$V(Z, Y) = E(G_\sigma(Z - Y))$$

$$= \iint_{x,y} G_\sigma(z - y) f_{ZY}(z, y) dz dy \quad (2)$$

Correntropy 作为自适应系统训练的一中新的代价函数,用于度量随机变量间的局部相似性。使用 Parzen 窗估计可以得到 Correntropy 的一种近似表达,求解其最大值就得到了 MCC:

$$\hat{v}(z, y) = \frac{1}{N} \sum_{i=1}^N G_\sigma(z_i - y_i) = \frac{1}{N} \sum_{i=1}^N G_\sigma(e_i)$$

$$MCC \Leftrightarrow \max \hat{v}(z, y) \quad (3)$$

其中 $e_i = z_i - y_i$ 。MCC 准则作为一种局部相似度准则,它对于样本中的一些奇异值具有鲁棒性。

1.2 最小均方误差(MMSE)准则

MMSE 作为误差评价准则是通过最小化误差分布的方差来实现的。其表达式如下:

$$MMSE(Z, Y) = E((Z - Y)^2)$$

$$= \iint_{x,y} (z - y)^2 f_{ZY}(z, y) dz dy$$

$$\Leftrightarrow \frac{1}{N} \sum_{i=1}^N (z_i - y_i)^2 \quad (4)$$

MMSE 仅仅包含了信号的二阶统计量,能够较好地处理高斯过程,但对于具有非高斯和非线性特性的随机过程来说,由于信号的统计特性需要更高阶的统计量来描述,并且 MMSE 对于奇异点的处理不具有鲁棒性,因此使用 MMSE 误差较大。

1.3 MCC 和 MMSE 的混合准则

MCC 与 MMSE 作为两种目标评价函数,分别在处理非线性与线性问题中有其优势,而且对于 MCC 选取特殊的窗函数就与 MMSE 一致,因此在这里将两者结合,构造一种混合的代价函数,如下式所示:

$$MCC_MMSE \Leftrightarrow \gamma \max \frac{1}{N} \sum_{i=1}^N G_\sigma(e_i)$$

$$+ (1 - \gamma) \min \frac{1}{N} \sum_{i=1}^N \frac{1}{2} e_i^2 \quad (5)$$

其中 $\gamma \in (0, 1)$ 为权重系数。由该式很容易看出当 $\gamma = 1$ 时,其退化为 MCC, $\gamma = 0$ 时即是 MMSE。

文献[7]对 correntropy 的概念进行了详细的讨论,证明其可作为衡量任意两个随机变量的距离度量,它满足距离度量的所有性质。使用 MCC 作为评价准则来训练自适应系统,可以使得系统的输出在该距离度量意义上更接近期望值。使用 MCC 作为一种衡量局部相似度的准则,其性能优于传统的 MMSE 准则,对于具有非零均值、非高斯特性且具有奇异值的数据处理非常有用。

2 基于 MCC 的 Elman 神经网络流量预测模型

利用神经网络通过学习理论进行数据建模,从而可逼近任意非线性函数的优点,且反馈式 Elman 神经网络具有输入延迟,所以将其应用于网络流量预测非常合适。

2.1 Elman 神经网络模型结构

Elman 神经网络是一种动态反馈神经网络^[12],其结构一般分为 4 层:输入层、隐含层、输出层和承接层。图 1 为 Elman 神经网络模型及其各层之间的表达式。如果设输入向量 u 为 r 维向量,输出向量 y 为 m 维,隐含层输出向量 x 为 n 维,承接层输出向量 x_c 为 n 维, w^1, w^2, w^3 分别为隐含层到输出层、输入层到隐含层、承接层到隐含层的连接权值。 $g(\cdot)$ 为输出神经元的激活函数, $v(\cdot)$ 是隐含层输出的线性组合, $f(\cdot)$ 为隐含层神经元的激活函数,可以采用线

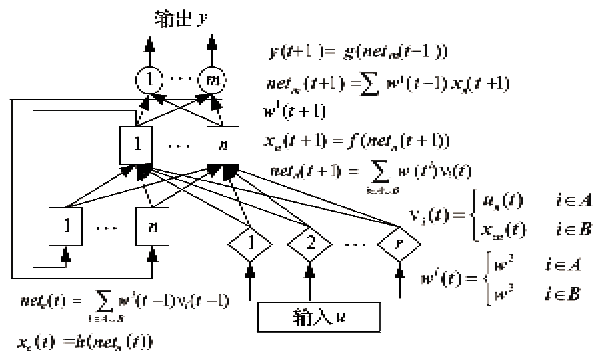


图 1 Elman 神经网络模型及各层之间的公式

性或非线性函数, $h(\cdot)$ 为承接层的函数, 承接层用来记忆隐含层前一时刻的输出值并返回给输入, 可认为是一步时延算子^[13], $net(\cdot)$ 表示某层的净输入, 用 A 表示输入层, 用 B 表示承接层, t 为迭代次序。

2.2 基于 MCC 的 Elman 神经网络的学习过程

Elman 神经网络采用 BP 算法进行权值的修正, 传统的学习指标函数都是基于 MMSE 的。这里将 MCC 作为 Elman 神经网络的学习指标函数, 给出基于 MCC 的 Elman 神经网络的学习算法, 图 2 为学习算法流程图。

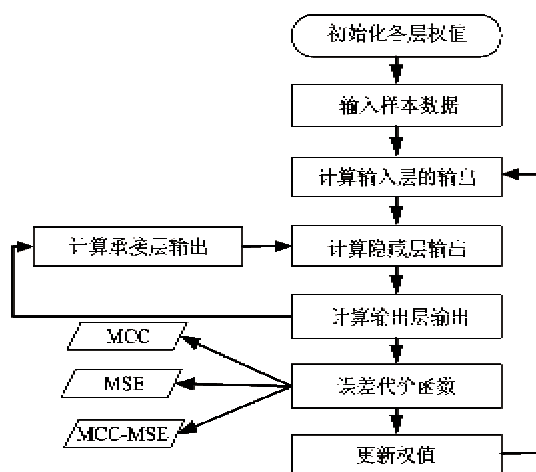


图 2 学习算法流程图

使用 MCC 作为 Elman 神经网络误差的评价函数, 如下式所示:

$$\text{cost}F_MCC \Leftrightarrow \frac{1}{N} \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^N e^{-\frac{(y_i^R - y_i^E)^2}{2\sigma^2}} \quad (6)$$

其中 y_i^R , $y_i^E = W_n^T X_i$ 分别为实际期望值和输出层网络输出预测值和, N 为样本数, W_n^T 为权值向量的转置, X_i 为第 i 组输入样本。寻找能够最大化代价函数 $\text{cost}F_MCC$ 的权值向量非常困难, 因此采用简单梯度上升算法更新权值, 如下式所示:

$$W(n+1) = W(n) + \eta \nabla \text{cost}F_MCC \quad (7)$$

通过计算 $\text{cost}F_MCC$ 的梯度可得

$$\begin{aligned} \nabla \text{cost}F_MCC = & \\ & \frac{1}{N} \frac{1}{\sqrt{2\pi}\sigma^3} \sum_{i=1}^N e^{-\frac{(y_i^R - W_n^T X_i)^2}{2\sigma^2}} (y_i^R - W_n^T X_i) X_i \end{aligned} \quad (8)$$

从而可以得到权值学习过程如下:

$$\begin{aligned} W(n+1) = W(n) + \frac{\eta}{N} \frac{1}{\sqrt{2\pi}\sigma^3} \sum_{i=1}^N e^{-\frac{(y_i^R - W_n^T X_i)^2}{2\sigma^2}} (y_i^R \\ - W_n^T X_i) X_i \end{aligned} \quad (9)$$

η 是学习率。

2.3 网络流量预测模型

(1) 数据采集与处理

为了验证基于 MCC 的网络流量预测模型的有效性, 以 <http://newsfeed.ntcu.net/~news/2006/> 流量库中 2006 年的前 150 天的实际流量数据来进行实验。图 3 所示为所采集的一年 365 天的网络流量数据图, 可看出网络流量具有突发性, 随机变化等特点。由于网络流量原始数据数量级差别和神经网络预测性能与输入输出数据数量级差别有很大的关系, 为了提高预测精度, 加快学习速度, 在网络训练

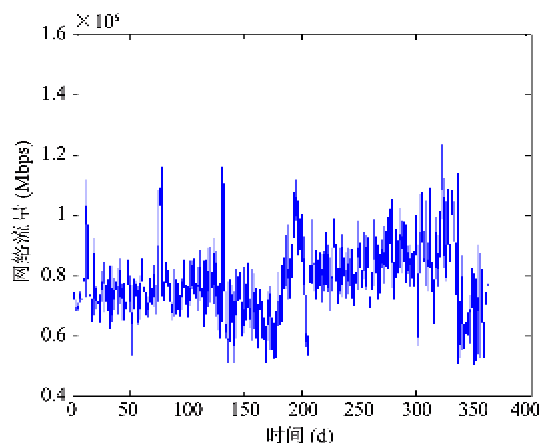


图 3 网络流量图

之前需先按下式对数据进行归一化处理:

$$d_{\text{norm}}(t) = \frac{d(t) - \min(d(t))}{\max(d(t)) - \min(d(t))} \quad (10)$$

其中 $d(t)$ 表示网络流量的原始数据, $d_{\text{norm}}(t)$ 表示归一化后的数据, $\max(d(t))$ 和 $\min(d(t))$ 分别表示原始流量数据的最大值和最小值。

(2) 网络流量预测模型

将归一化后的数据分为两组, 一组作为训练样本数据, 一组作为测试数据。

在 2.2 中描述了基于 MCC 的 Elman 神经网络的学习过程, 在此对于使用 MCC 评价准则的 Elman 神经网络用于网络流量预测的算法进行描述:

步骤 1: 网络初始化: 选择网络输入输出维数、隐藏层节点个数、初始权值、滑动窗口大小和阈值。

步骤 2: 网络训练: 根据 2.2 中描述的学习过程, 以 MCC 为评价准则, 根据历史样本流量数据对神经网络权值进行训练。

步骤 3: 当学习过程收敛到预定的阈值时, 停止学习过程, 此时神经网络权值确定。

步骤 4: 根据训练好的网络预测未来的网络流量数据。

3 网络流量预测仿真与分析

使用神经网络进行网络流量的预测,首先需要确定网络的输入,输出节点,使其反映网络系统流量的规律。文中以一周 7 天的数据作为网络的输入,输出为第 8 天的流量数据,也即输入节点个数与输出节点个数分别为 7 和 8。

以 2 中的 3 个准则作为 Elman 神经网络学习过程中的目标函数,使用训练好的网络进行流量预测。

3.1 网络流量特性分析

文献[9]对网络流量的特性进行了研究,发现在大多数情况下,网络流量呈现非高斯性和非线性特征,这是本文使用 MCC 作为相似度评价准则的基础。峰度 f 是一种判定信号是否具有非高斯特性的概念,若 $f = 0$, 则信号呈现出高斯性,否则表明其具有非高斯性。为了进一步验证网络流量的非高斯特性,以峰度作为网络流量特性研究的工具。网络流量峰度 f 的定义为 $f = \frac{mean(D^4 - 3mean[D^2])}{mean[D^2]^2}$, 其中 D 为样本数据集, $mean()$ 表示均值。将样本数据带入上式,计算峰度为 1.1275,可见该流量数据的确具有非高斯特性,并从其数据图

3 可看出网络流量具有明显的突变特征。

3.2 网络流量预测结果比较分析

为了说明将 MCC 和 MCC-MMSE 准则用于 Elman 神经网络进行网络流量预测的有效性,从隐含层节点个数、核大小、参数 γ 3 个方面对上述 3 个评价准则进行的仿真比较分析。

(1) 不同隐含层节点个数下的预测效果比较

合理确定 Elman 神经网络的结构是预测性能优劣的基础。隐藏层节点数是影响网络收敛速度和预测精确度的一个因素^[12]。过多的网络节点会增加训练网络的时间且会使网络的泛化能力减弱,网络的预测能力下降。但网络节点过少则建模不充分。在隐含层中神经元数目的确定是一个经验性的问题,需要大量的实验。下式确定三层网络中隐含层神经元个数:

$$n_2 = 2n_1 + 1 \quad (11)$$

其中 n_2 为隐含层神经元个数, n_1 为输入层神经元个数。在这里取隐含层神经元个数分别为 7、11、14、18 来进行网络流量预测。设 MCC 和 MCC-MMSE 准则中核大小 $\sigma = 0.53, \gamma = 0.45$ 。图 4 为不同隐含层节点个数下三种评价准则的网络流量预测结

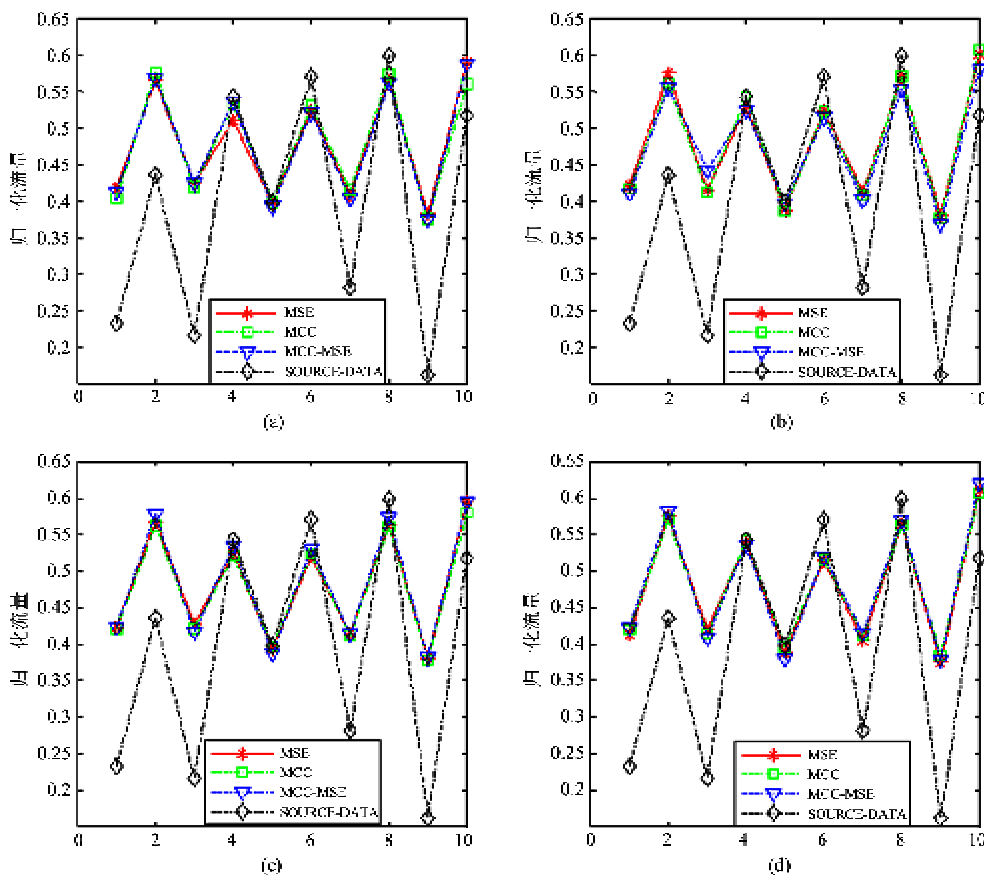


图 4 预测结果比较图

果。由图可知,基于 MCC 的预测结果更接近于原始值,而 MMSE 和 MCC-MMSE 的结果相对来说不如 MCC 的效果好,由于 MCC 从信息熵的角度考虑到了网络流量数据的非高斯特性,更适合于流量预测,实验结果也达到了预期的设想。

图 5 为相应的预测误差结果图,虽然在隐含层神经元的个数不同,但可以看出 MCC 的平均误差还是小于 MMSE 和 MCC-MMSE 的误差。并且在隐含层个数为 14 时,三者的预测误差都达到了最小。

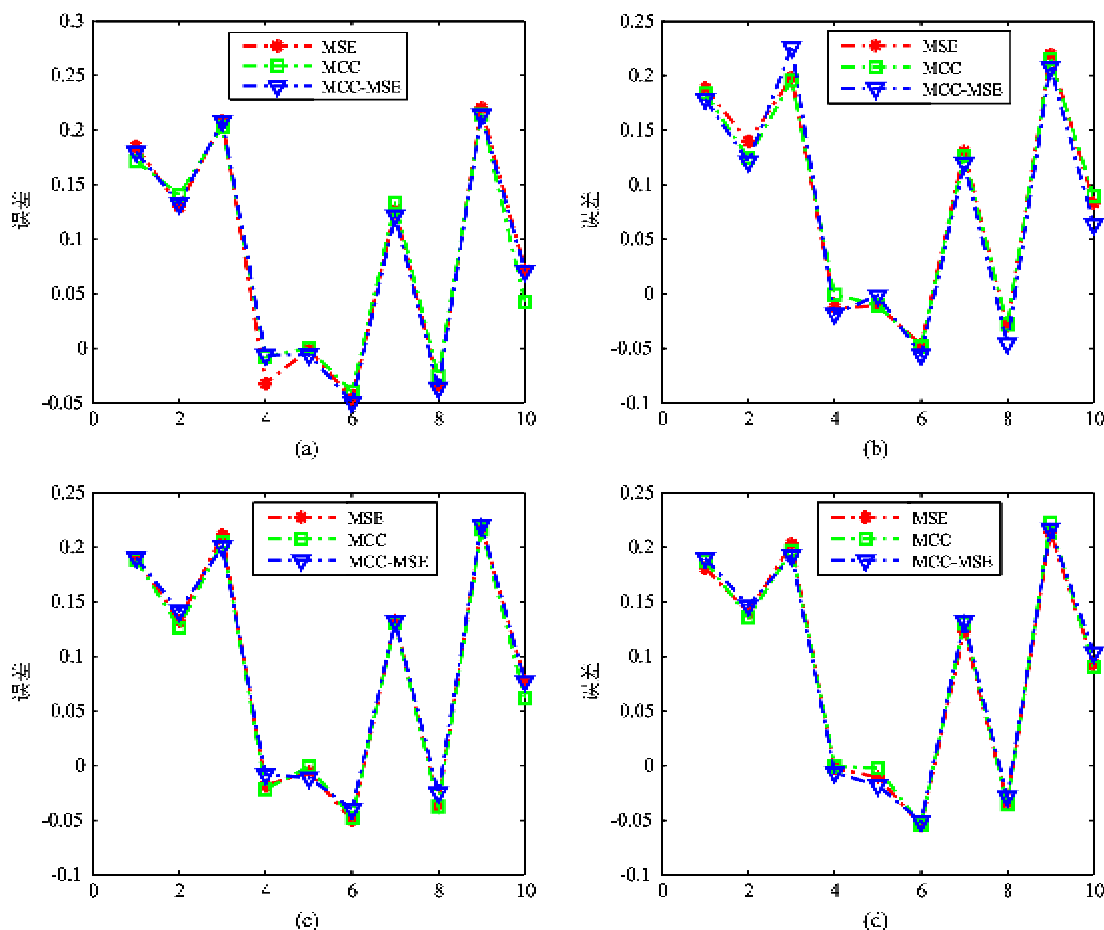


图 5 预测误差结果图

均值绝对误差 (mean absolute error, MAE)、均值相对误差 (mean relative error, MRE) 及和方差 (sum squared error, SSE) 是广泛应用于网络预测结果优劣的评价准则,在此根据这 3 种准则对 3 种评价准则下的预测结果进行比较。在计算过程中基于 4 个隐藏层个数的平均后求得 MAE、MRE 和 SSE。表 1 给出了不同准则预测结果的评价,从中可知 MCC 的结果优于其他两者。MCC 使用高斯核函数,

它包含了误差的高阶统计信息,因此可以提高误差的精度。而 MMSE 只包含二阶统计信息,因此相对于 MCC 来说,其预测结果的误差精度要高。

(2) 隐藏层数 ($N_{\text{hide-node}}$) 为 11, 不同核大小 σ 下的比较

MCC 包含核参数,该参数的选取对网络的学习收敛性能具有一定的影响。对于核大小的选择,文献中提出可以通过 sliver-man 准则确定核大小,也可以根据具体的问题人为设定,为研究核大小对预测结果的影响,分别取核大小为: 0.22、0.37、0.52、0.67、0.95 进行预测。图 6 为不同 σ 下的 MCC 与 MCC-MMSE 的预测结果误差曲线图。

如图 6 所示,选取不同 σ , MCC 的预测误差相对小于 MCC-MMSE 的结果,可见在混合准则中,MMSE 对误差的精度具有一定的影响。表 2 以 MAE

表 1 不同准则预测结果的评价

误差评价准则	MAE	MRE	SSE
MCC	0.1024	0.4102	0.6570
MCC-MMSE	0.1048	0.4176	0.6726
MMSE	0.1061	0.4217	0.6811

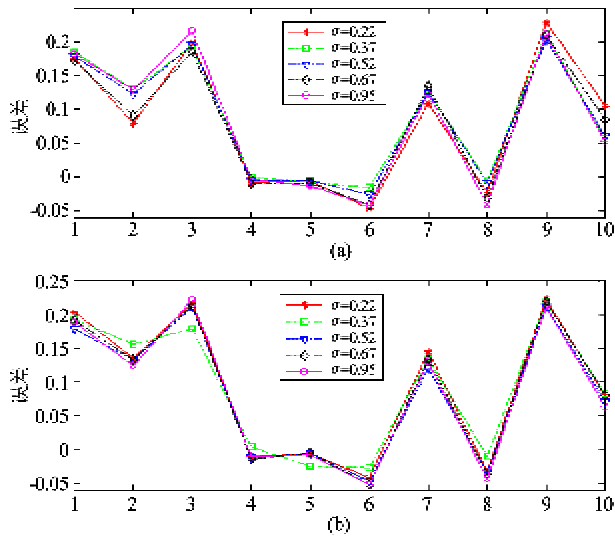


图 6 不同 σ 下 MCC 与 MCC-MMSE 预测结果误差比较 ($N_{\text{hide-node}} = 11$)

表 2 MAE 结果

核 σ	0.22	0.95	0.37	0.67	0.52
评价准则					
MCC	0.0982	0.1014	0.0933	0.0972	0.0930
MCC-MMSE	0.1094	0.1052	0.1030	0.1080	0.1023

为性能评价准则,给出了不同核大小下的两种准则预测结果的 MAE 值,可见 MCC 的 MAE 值要低于 MCC-MMSE 的值,且在核大小为 0.52 时, MCC 的误差最小,预测精度有很大的提高。

(3) 针对 MCC-MMSE,对于不同 γ 的比较

在 MCC-MMSE 准则中引入了权重因子,为了分析其对预测结果的影响,这里分别取值:0.2, 0.4, 0.6, 0.8 进行试验。图 7 (a) 和图 7 (b) 分别为预测

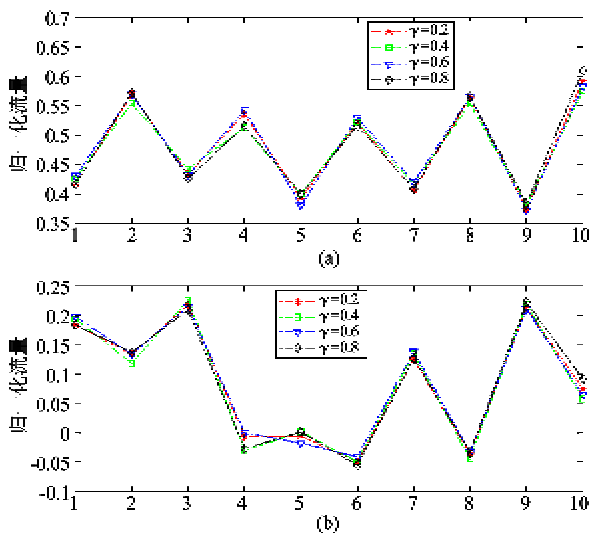


图 7 不同 γ 的 MCC-MMSE 的预测结果与误差图

结果与误差曲线图,从该图中可看出其预测结果随着权重的增加,预测精度有所改进,因为当权重增加时,该准则中 MMSE 的权重将有所减小,更大地体现了 MCC 的优越性,因此其预测误差将会减小。表 3 给出了使用 MCC-MMSE 准则进行预测时,选取不同 γ 所得到的 MAE 的值,从中也可得出上述变化趋势。

表 3 MCC-MMSE 预测结果 MAE(不同 γ)

γ	0.2	0.4	0.6	0.8
MAE	0.1088	0.1052	0.1076	0.1044

4 结论

网络流量预测误差与流量的非高斯误差概率分布相关。已有的网络流量预测研究都以 MMSE 作为神经网络训练模型的误差标准函数,它只包含数据的二阶统计信息,因此在网络流量预测过程中的误差较高,并且使用 MMSE 准则的自适应系统对于非线性问题的处理具有一定的缺陷。本文通过对最大相关熵准则(MCC)进行分析研究发现,该准则可以很好地处理如网络流量这种具有非高斯、非线性特性的数据问题,在此基础上建立了基于 MCC 和 MCC-MMSE 的 Elman 神经网络预测模型,并用于网络流量预测,仿真结果表明其性能优于传统的 MMSE 准则。在后续的研究中,需要根据网络流量采集工具进行现有网络中实际流量的采集,应用该模型进行流量预测,验证其可行性。

参考文献

- [1] Miguel I G, Jose R G, Luis O B, et al. On the dynamic allocation of resources using linear prediction of aggregate network traffic. *Computer Communications*, 2003, 26 (3): 1341-1352
- [2] Garroppo R G, Gianlano S, Pagano M, et al. On traffic prediction for resource allocation: a Chebyshev bound based allocation scheme. *Computer Communications*, 2008, 31(2): 3741-3751
- [3] Maysam M, Abdallah S. Traffic-prediction-assisted dynamic bandwidth assignment for hybrid optical wireless networks. *Computer Networks*, 2012, 56(1): 244-259
- [4] 姜明,吴春明,张大明. 网络流量预测中的时间序列模型比较研究. *电子学报*, 2009, 37(11): 2353-2357
- [5] Hussein D. An object-oriented neural network approach to short-term traffic forecasting. *European Journal of Opera-*

- tional Research*, 2001,131(2):253-261
- [6] Eleni I V, Matthew G K, John C G. Optimized and meta-optimized neural networks for short-term traffic flow prediction: a genetic approach. *Transportation Research Part C: Emerging Technologies*, 2005,13(3):211-234
- [7] Park D C. Prediction of MPEG video traffic over ATM networks using dynamic bilinear recurrent neural network. *Applied Mathematics and Computation*, 2008, 205(2): 648-657
- [8] Zhu X W, Li F R. Traffic flow prediction based on artificial life and RBF neural network. *Energy Procedia*, 2011, 10(3):1250-1254
- [9] Kolbusz J, Paszczyński S, Wilamowski B M. Network Traffic Model for Industrial Environment. *IEEE Transactions on Industrial Informatics*,2006,2(4):213-220
- [10] Liu W F, Pokharel P, Principe J C. Correntropy: a localized similarity measure. In: Proceedings of the International Joint Conference on Neural Network, Vancouver, Canada, 2006. 179-184
- [11] Liu W F, Pokharel P, Principe J C. Correntropy: properties and applications in non-Gaussian signal processing. *IEEE Transactions on Signal Process*, 2007, 55(11): 5286-5298
- [12] Shi X H. Improved Elman networks and applications for controlling ultrasonic motors. *Applied Artificial Intelligence*, 2004,18(7):603-629
- [13] Liang X R, Fan Y K, Li J Y. Freeway traffic flow model based on rough set s and Elman neural network. In: Proceedings of the International Conference on Information Engineering and Computer Science, Wuhan, China, 2009.1-4

Network traffic prediction based on maximum correntropy criterion

Qu Hua^{*}, Ma Wentao^{*}, Zhao Jihong^{*}, Wang Tao^{**}

(^{*} School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049)

(^{**} School of Telecommunication and information Engineering,
Xi'an University of Posts and Telecommunications, Xi'an 710121)

Abstract

With the nonlinear characteristics of network traffic considered, a new network traffic prediction method based on the maximum correntropy criterion (MCC), a new error evaluation criterion, was proposed to improve the precision of traffic network prediction. The method uses the MCC to train Elman neural networks, and this evaluation criterion is based on the new concept of a new similarity function, the generalized correlation entropy (correntropy) function, which takes the Parzen window estimation of the error probability density function and the Rayleigh entropy as the basis. Simultaneously, a mixed evaluation criterion which combines the MCC and the minimum mean square error (MMSE) criterion was presented. In view of the characteristics of traffic networks such as the nonlinear, non Gauss, and mutation, the Elman neural network was trained by the MCC and the mixed criterion, respectively, and then a trained neural network was used to predict network traffic. The simulation results show that the accuracy of the prediction is superior to the prediction results of the Elman neural network with the MMSE criterion.

Key words: maximum correntropy criterion (MCC), minimum mean square error (MMSE), Elman neural network, traffic network prediction