

# 基于 EMD 聚类的实时网络流量预测模型

姚立霜 刘丹 裴作飞 王云锋

重庆邮电大学通信与信息工程学院 重庆 400065

**摘要** 针对复杂的网络流量呈现出的多种特性,传统的单一模型预测效果差。为了提高流量预测的准确性和实时性,提出了一种基于经验模态分解(EMD)和聚类的网络流量预测模型。首先通过 EMD 将网络流量分解为不同时间尺度上频率单一的本征模函数(IMFs);其次通过改进的 K 均值聚类算法对 IMF 分量做聚类分析,将复杂度相近的 IMF 分量聚到一起;然后对聚类的 IMF 分量用自回归移动平均(ARMA)模型进行预测;最后将各 IMF 分量序列的预测值进行求和得到网络流量的预测值。实验结果证明,与 EMD-ARMA 模型相比,该模型不仅缩短了训练耗时,且均方误差(MSE)、平均绝对误差(MAE)分别下降了 13.8%和 7.6%,趋势预测准确率(APT)提高了 6%,提高了网络流量的预测精度,可用于实时流量预测。

**关键词** 网络流量;经验模态分解;K 均值聚类;自回归移动平均;流量预测

**中图分类号** TP393

## Real-time Network Traffic Prediction Model Based on EMD and Clustering

YAO Li-shuang, LIU Dan, PEI Zuo-fei and WANG Yun-feng

School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

**Abstract** Based on the multiple characteristics of complex network traffic, the traditional single model has poor prediction results. In order to improve the accuracy and real-time performance of traffic prediction, a network traffic prediction model based on EMD and clustering is proposed. First, the network traffic is decomposed into IMFs through EMD. IMFs are on different time scales and their frequencies are relatively single. Secondly, IMFs are clustered by an improved K-means clustering algorithm, and IMFs with similar complexity are gathered. Then the clustered IMFs are predicted using the ARMA model. Finally, the predicted values of each IMF are summed to obtain the predicted value of overall network traffic. Experimental results show that, compared with the EMD-ARMA model, the model not only reduces the training time, and its MSE and MAE reduce by 3.8% and 7.6% respectively, APT improves by 6 percentage. The model achieves higher prediction accuracy of network traffic and can be used for real-time traffic prediction.

**Keywords** Network traffic, EMD, K-means clustering, ARMA, Traffic prediction

## 1 引言

随着 5G 时代的到来,网络流量和上网人数急剧增长。复杂的网络互联环境,造成了网络的安全性与稳定性较低,从而影响了网络服务的质量。为了有效地维护网络,提高网络性能,网络管理人员需要实时监控网络,通过网络流量的预测值来判断网络发生异常的可能性,以在网络拥塞之前采取行动来解决网络问题,从而提供更好的服务给用户<sup>[1]</sup>。因此,网络流量预测在网络管理和维护中起着十分重要的作用。

现有的网络流量预测模型主要有两类:线性预测和非线性预测。随着互联网应用和流量的增多,实际的网络流量表现出了自相似、多尺度、非线性等多种特性<sup>[2]</sup>,直接采用自回归移动平均(Autoregressive moving average model, ARMA)<sup>[3]</sup>或差分整合移动平均自回归(Autoregressive Integrated Moving Average, ARIMA)<sup>[4]</sup>等线性预测模型都难免造成较大的预测误差。神经网络<sup>[5]</sup>作为非线性预测模型,具有逼近任何非线性函数的能力,其预测精度有一定的提高,但同时

也会导致模型复杂度和计算复杂度的增加。面对复杂的网络流量特性,如何构建合适、准确的网络流量预测模型一直是网络管理领域的研究重点。

为了准确描述网络流量的复杂特性,需要对其在多尺度下进行分析。具有多分辨率分析能力的小波变换<sup>[6]</sup>是处理非平稳时间序列的有效方法之一。文献<sup>[7]</sup>对流量序列进行小波分解,对重构后的近似部分和各细节部分别建立自回归模型,进而实现原始网络流量的预测,比传统的预测方法具有更高的预测精度。但小波变换仍有许多不足之处需要克服,一方面,不同领域使用需要不同的小波函数,另一方面,一旦确定了分解尺度,小波变换的结果就是一定频带下的信号,不能改变。换句话说,小波变换本质上不是一种自适应信号处理方法<sup>[8]</sup>。作为一种时频信号处理技术,经验模态分解(Empirical Mode Decomposition, EMD)被提出来处理非线性和非平稳问题。EMD 在不需要预先确定参数的情况下,通过筛选过程将时间序列自适应分解成一系列的有模态函数(Intrinsic Mode Function, IMF)<sup>[9]</sup>。这些 IMFs 突出了原始信号不同

基金项目:长江学者和创新团队发展计划(IRT\_16R72)

This work was supported by the Changjiang Scholars and Innovative Research Team Program in University (IRT\_16R72).

通信作者:姚立霜(1096315681@qq.com)

的局部特征信息,并且每个IMF之间是相互独立的。因此,EMD解决了小波变换中分解尺度和小波函数的选择问题,已经被广泛用于流量预测之中。

文献[10]用EMD代替小波变换,将网络流量自适应分解成多个相对平稳的IMFs,再用短相关的ARMA模型预测各个IMF分量。但是原序列的非平稳性越强,经EMD分解得到的IMF个数就会越多,如果直接对每个IMF分量进行预测,这就需要较多的预测器,这样会增加计算的规模。

针对现有预测模型的不足,提出一种基于EMD聚类的实时网络流量预测模型。首先通过EMD分解将网络流量分解为不同时间尺度上、频率更加单一的IMF分量;其次通过改进的K均值聚类算法对IMF分量做聚类分析,将复杂度相近的IMF分量聚到一起;然后采用ARMA模型对聚类后的IMF分量进行建模和预测;从而减少预测器的个数,减小系统复杂度。通过对真实数据集的实验证明,该模型能有效提高实时网络流量的预测精度。

## 2 基于EMD聚类的实时网络流量预测模型

针对流量预测的准确性和实时性,提出了一种基于EMD聚类的实时网络流量预测模型。如图1所示,该模型分为5个部分:数据预处理、模态分解、分量聚类、分量预测、模态重构。

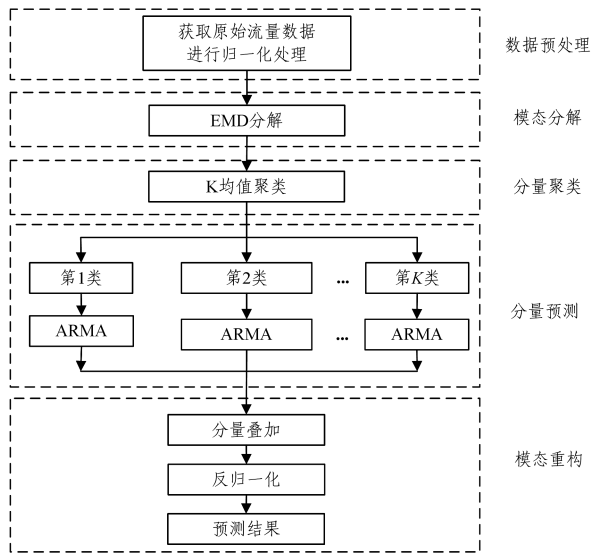


图1 基于EMD聚类的实时网络流量预测模型

Fig.1 Real-time network traffic prediction model based on EMD and clustering

第一部分是数据预处理,获取原始流量数据并进行预处理;为了加快计算速度,采用归一化处理:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

其中, $x_{\max}$ 和 $x_{\min}$ 分别是流量最大值和流量最小值。第二部分是模态分解,对预处理后的流量序列进行EMD分解,将复杂的网络流量序列分解为不同时间尺度上、成分简单、相对平稳的IMFs,以降低非线性、非平稳流量序列的预测偏差。第三部分是分量聚类,采用改进的K均值算法对分解后所得的子模态进行聚类分析,将复杂度相近的分量聚到一起,形成新的子序列。第四部分是分量预测,采用ARMA模型分别对新的子序列进行预测。第五部分是模态重构,将各分子序列的预测结果求和,并按式(2)进行反归一化处理,从而得到最

终的整体流量预测值。

$$x = x'(x_{\max} - x_{\min}) + x_{\min} \quad (2)$$

### 2.1 EMD分解

网络流量序列本身就是非线性、非平稳的时间序列,因此第二部分的模态分解采用EMD算法,将时间序列自适应分解为一组IMFs。将EMD算法应用于给定的流量序列 $x(t)$ 时,提取IMFs的基本过程如下<sup>[11]</sup>:

Step1 确定 $x(t)$ 的所有极大值点,利用三次样条插值得到相应的上包络线,同理,根据 $x(t)$ 的极小值可以得到下包络线;

Step2 根据上、下包络线得到均值 $m(t)$ :

$$m(t) = (e_{\min}(t) + e_{\max}(t)) / 2 \quad (3)$$

Step3 根据信号 $x(t)$ 和均值 $m(t)$ 提取细节信号 $h(t)$ :

$$h(t) = x(t) - m(t) \quad (4)$$

Step4 若 $h(t)$ 符合IMF,就是第一个模态分量IMF1(记为 $c_1$ );否则令 $h(t) = x(t)$ ,跳到Step1继续筛选;

Step5 余项 $r_1 = x(t) - c_1$ ,令 $r_1 = x(t)$ ,跳到Step1继续筛选,可以得到第二个模态分量IMF2,以此类推,得到相应的余项 $r_1 - c_2 = r_2, \dots, r_{n-1} - c_n = r_n$ 。当 $r(t)$ 是单调序列时,说明不能再分解出新的IMF,整个筛选过程就结束了。

通过上述筛选过程,最后流量序列 $x(t)$ 可分解为IMFs和余项:

$$x(t) = \sum_{i=1}^n c_i(t) + r(t) \quad (5)$$

### 2.2 聚类分析

第三部分的分量聚类指将EMD分解得到所有的IMF分量按一定的方式划分成若干类,形成新的子序列,从而挖掘出IMF内部之间的隐藏联系,减少预测器的个数。

原始K均值算法<sup>[12]</sup>是聚类分析中最常用的算法,其算法简单、计算快速,在各个领域都有着广泛的应用。因为原始K均值算法对孤立点数据敏感,且初始聚类中心的不同会导致聚类结果的不同,因此孤立点数据和随机选取的初始聚类中心会大大影响K均值算法的聚类结果。本文使用改进的K均值算法对分量进行聚类分析。步骤如下:

Step1 计算每个IMF分量与原始流量序列的相关系数,剔除相关系数最小的干扰分量。

$$\text{Corr}(x, y) = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (6)$$

Step2 基于样本熵(Sample Entropy, SampEn)选择K个聚类中心,使各聚类中心的距离尽量远。计算每个IMF分量的SampEn值,将SampEn最大的IMF分量作为第一个聚类中心Z1,将SampEn最小的IMF分量作为第二个聚类中心Z2,将SampEn中值的IMF分量作为第三个聚类中心Z3。K值一般取2或3即可。其中SampEn<sup>[13]</sup>的估计公式为:

$$\text{SampEn}(m, r, N) = -\ln(B_{m+1}^r / B_m^r) \quad (7)$$

其中, $m$ 为维数,取值1或2; $r$ 为相似容限,取值范围为 $[0.1std \sim 0.25std]$ , $std$ 为标准偏差; $N$ 为原始时间序列的长度; $B_m^r(i)$ 和 $B_{m+1}^r$ 分别为两个序列在相似容限 $r$ 下匹配 $m$ 和 $m+1$ 个点的概率。

Step3 计算其他IMF分量序列与K个聚类中心的距离,距离值越小说明它们之间的内在联系越大,因此将各IMF分量分配到距离其最近的聚类中心所在的簇,距离公式如下:

$$d(x, y) = \sqrt{\sum_{i=1}^N (x_i - y_i)^2} \quad (8)$$

Step4 计算每一个簇所有样本的均值,将均值作为该簇新的聚类中心。

Step5 重复 Step3 和 Step4,直至聚类中心不再改变,整个聚类过程就结束了。

### 2.3 ARMA 模型

为了达到实时性的效果,因此第四部分的分量预测必须选择模型简单、计算快速的预测模型。作为线性预测的代表,ARMA 模型通过多项式设置足够多的参数拟合历史数据来预测未来数据,具有模型简单、实现容易、计算速度快等优势,适合预测平稳的时间序列,短期预测精度较高<sup>[14]</sup>。模型表示为:

$$X(t) = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (9)$$

其中,  $\varepsilon_t$  为白噪声序列;  $X(t)$  为平稳时间序列。

ARMA 模型的建模步骤如下:

Step1 通过赤池信息量准则(Akaike's Information Criterion, AIC)来确定模型的最佳阶数,当  $A(s)$  达到最小时,此时的  $p, q$  就是最佳的阶数。

$$A(s) = \ln \hat{\sigma}^2 + \frac{2(p+q+1)}{N} \quad (10)$$

其中,  $\hat{\sigma}^2$  为  $\varepsilon_t$  的方差;  $N$  为流量序列长度。

Step2 采用最小二乘法估计未知参数的值。记未知参数为:

$$\tilde{\beta} = (\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q) \quad (11)$$

$$F_t(\tilde{\beta}) = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (12)$$

$$Q(\tilde{\beta}) = \sum_{t=1}^n \varepsilon_t^2 = \sum_{t=1}^n (x_t - F_t(\tilde{\beta}))^2 = \sum_{t=1}^n [(x_t - \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \dots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q})^2] \quad (13)$$

按式(13)计算残差项平方和  $Q(\tilde{\beta})$ , 分别求  $Q(\tilde{\beta})$  对待估计参数的偏导数,并令其偏导数为零。通过求解方程组,找到  $Q(\tilde{\beta})$  的极值点。此时的参数值  $\varphi_1, \varphi_2, \dots, \varphi_p, \theta_1, \theta_2, \dots, \theta_q$  就是最小二乘法的最优解,也就是所求的估计值。

Step3 检验残差序列是否为白噪声序列。若通过检验,说明没有信息需要进一步提取,则建模完成,否则返回 step1。

Step4 根据估计好的参数和式(9)就可以建立相应的 ARMA 模型。并利用建立好的模型,根据前  $m$  个时刻的流量值可预测得到第  $m+1$  时刻的流量值,实现短期预测。

## 3 实验仿真

### 3.1 数据来源

本文所用的流量数据来自日本 MAWI 工作组<sup>1)</sup>,从 2020 年 1 月 1 日 14:00 至 14:15,累积形成流量时间序列,取其中 800 个样本点用于训练,50 个数据用于测试。训练数据经归一化处理得到的波形如图 2 所示,该流量序列的波动起伏很大,具有明显的非线性、非平稳性。归一化处理不影响流量特性,只是把流量数据值映射到了 0~1 范围之内,方便后续的计算,提升模型的收敛速度。

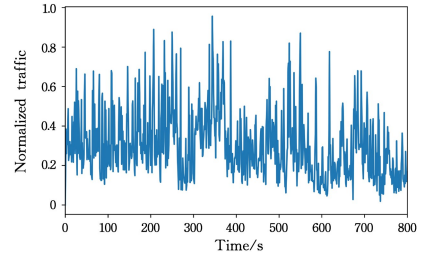


图 2 归一化流量序列

Fig. 2 Normalized traffic sequence

本文所有仿真均在 16.00 GB, 64 bit, Windows 10, Intel Core i7-6100@3.70 GHz 的计算机系统上通过 Ipython Notebook 编程实现。

### 3.2 评价指标

本文采用 3 个评价指标,它们的定义分别如式(13)~式(15)所示,均方误差(MSE)、平均绝对误差(MAE)和趋势预测准确率(APT),从不同的方面定量地衡量模型的预测效果。MSE 和 MAE 反映了真实值与预测值之间的误差大小,APT 反映了其预测序列趋势的能力,因此 MSE 和 MAE 的值越小,APT 的值就越大,说明预测模型描述流量数据具有更好的精确度。

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - x_i')^2 \quad (14)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i - x_i'| \quad (15)$$

$$APT = \frac{1}{N} \sum_{i=1}^N T_i \quad (16)$$

其中,  $x_i$  和  $x_i'$  分别为网络流量在第  $i$  时刻的实际值和预测值;当  $(x_i - x_{i-1}) * (x_i' - x_{i-1}') \geq 0$  时,  $T_i = 1$ , 否则  $T_i = 0$ 。

### 3.3 实验分析

原始网络流量数据经过 EMD 分解后得到 7 个不同尺度下频率单一的 IMF 分量(IMF1~IMF7)和一个余项(residue)。分解结果如图 3 所示,从 IMF1 到 IMF7,再到 residue,其频率成分是从高频到低频依次分布的,随机性和突发性逐渐减弱,成分越来越简单,IMF 分量也越来越平滑。

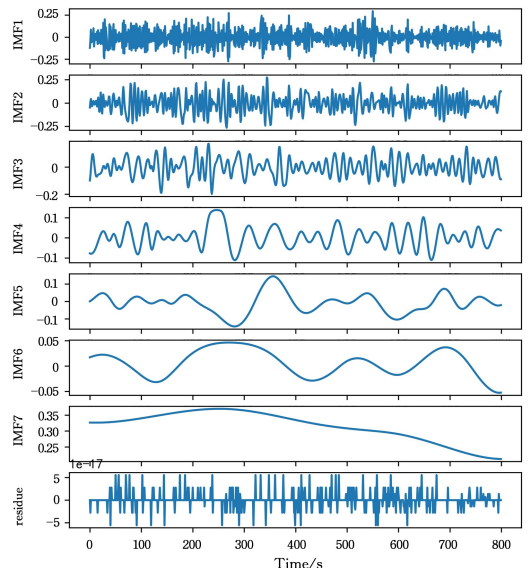


图 3 EMD 的分解结果

Fig. 3 Decomposition results of EMD

<sup>1)</sup> <http://mawi.wide.ad.jp/mawi>

计算每个分量与原始信号的相关系数,如图 4 所示,可以看出,余项与原始信号的相关系数最小,几乎为零,说明余项与原始信号的相关度极低,含有的有用信息极少,可以剔除,从而使聚类分量更能反映正常数据的特征。

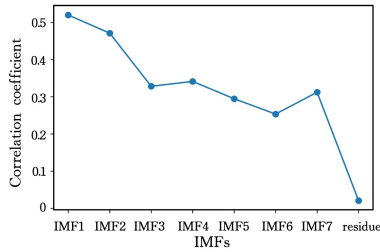


图 4 IMF 分量与原始信号的相关系数

Fig. 4 Correlation coefficient between IMFs and original signal

计算每个 IMF 分量的样本熵值,取  $m=2, r=0.2std$ ,得到的结果如图 5 所示。随着 IMF 分量频率的降低,各 IMF 的样本熵值是逐渐减小的,这就说明了各 IMF 分量序列产生新信息的可能性在逐渐减小,其序列的复杂度逐渐降低。

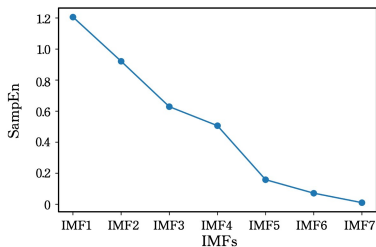
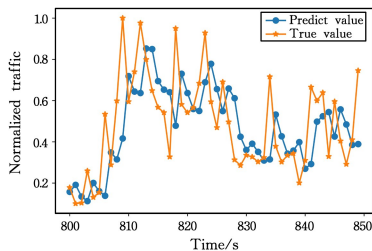


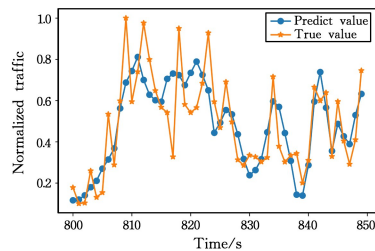
图 5 IMF 分量的样本熵值

Fig. 5 SampEn of IMFs

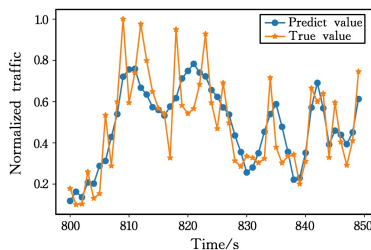
通过改进的  $K$  均值聚类算法,将复杂度相近的 IMF 分量分为一类,形成新的子序列,最终 8 个 IMF 分量被聚为 3 类,形成 3 个新的子序列。接着,通过 AIC 准则得到 3 个新



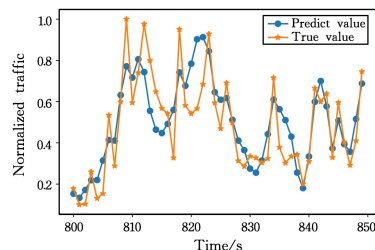
(a) ARIMA 模型的预测效果



(b) EMD-ARMA 模型的预测效果



(c) EMD-Elman 模型的预测效果



(d) 本文模型的预测效果

图 7 不同模型的预测效果

Fig. 7 Prediction effect of different models

进一步地,为了定量检测不同模型的预测效果,分别对这 5 种模型重复仿真 10 次,并对预测结果 MSE、MAE、APT 以及耗时分别求平均值,得到不同模型的预测评价指标如表 2 所列。

序列的阶数( $p, q$ ),具体的聚类结果和对应的阶数如表 1 所列,聚类后的效果如图 6 所示。

表 1 IMF 的聚类结果

Table 1 Clustering results of IMFs

New serial number	1	2	3
Original serial number	1	2, 3, 4	5, 6, 7
Order( $p, q$ )	(4, 1)	(4, 2)	(1, 3)

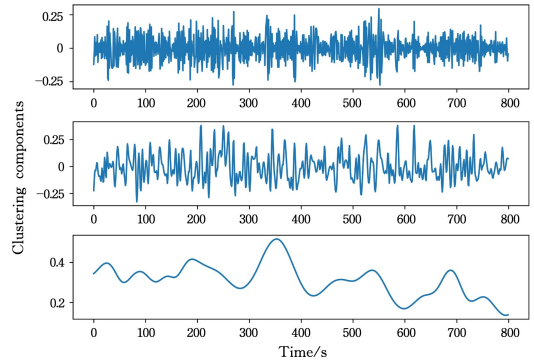


图 6 IMF 的聚类效果

Fig. 6 Clustering effect of IMFs

接着采用 ARMA 模型对这 3 个新的子序列进行预测。最后将 3 个新序列的预测结果相加,得到最终的流量预测值,再与实际的流量值进行比较。除了本文模型,还采取了 3 个对比模型,即 ARIMA 模型<sup>[14]</sup>、EMD-ARMA 模型<sup>[10]</sup>、EMD-Elman 模型<sup>[15]</sup>进行对比,其中涉及的 ARMA 模型均采用 AIC 准则确定各自阶数,即  $p$  和  $q$  值;Elman 神经网络通过经验公式确定网络结构为 50-10-1,迭代次数为 1000。4 个预测模型的实际值与预测值的拟合曲线如图 7 所示,可以看出本节所提模型表现出了更好的拟合效果,其各个预测值更接近实际的网络流量值。

与单一模型 ARIMA 的预测结果相比,本文模型的 MES、MAE 分别下降了 51.3% 和 29.1%,且 APT 有了大幅度提高,提升了 28%。虽然 ARIMA 模型的应用范围比 ARMA 模型更广,但也只能针对齐次的非平稳过程,

仍然具有很大的局限性。

与 EMD-ARMA 模型的预测结果相比,本文模型 MSE、MAE 分别下降了 13.8% 和 7.6%,且 APT 提高了 6%。这是因为本文方法通过改进的  $K$  均值聚类将复杂度相近的 IMF 聚到了一起,挖掘出 IMF 之间的隐藏信息。

与 EMD-Elman 模型的预测结果相比,本文模型 MES、MAE 还分别下降了 25.3% 和 15.8%,且 APT 提升了 4%。EMD-Elman 模型通过剔除 EMD 分解得到的高频 IMF 分量重构信号,减小噪声的干扰,再用 Elman 神经网络预测重构后的信号,其本质还是单一预测模型,其预测精度有限。

表 2 不同模型的综合评价指标

Table 2 Comprehensive evaluation index of different models

Method of prediction	MSE	MAE	APT/%
ARIMA	0.02928	0.13467	42
EMD-ARMA	0.01654	0.10344	64
EMD-Elman	0.01908	0.11337	60
Method of this article	0.01425	0.09550	70

表 3 列出了不同模型的训练耗时和预测耗时的对比,这是 4 种用于实时网络流量预测的模型,其预测耗时都控制在 1s 以内,但各自的训练耗时却有所差异。ARIMA 模型是单一线性预测模型,计算简单,因此其训练耗时最小。EMD-ARMA 模型对每个 IMF 分量进行预测,其计算复杂度增加,训练耗时也会较大。EMD-Elman 模型中含有神经网络,其收敛速度慢,训练过程需要大量的时间,导致它的训练耗时较大。而本文模型采用改进的  $K$  均值算法减少了预测器的数量,相比 EMD-ARMA 模型大大减小了计算复杂度,训练耗时也大幅度缩短。

表 3 不同模型的耗时对比

Table 3 Time-consuming comparison of different models

Method of prediction	Training time	Prediction time
ARIMA	0.30	0.76
EMD-ARMA	147.32	0.90
EMD-Elman	95.80	0.78
Method of this article	41.78	0.92

**结束语** 本文提出了一种基于 EMD 聚类的实时网络流量预测模型。首先通过 EMD 分解将网络流量分解为不同时间尺度上、频率更加单一、更容易预测的 IMF 分量;其次通过改进的  $K$  均值聚类算法将复杂度相近的 IMF 分量聚到一起;然后对聚类的 IMF 分量采用擅长线性预测的 ARMA 模型进行建模和预测,从而减少预测器的个数,减小整个系统的计算复杂度。实验仿真证明,本文模型在流量预测方面具有较好的准确性和实时性,将预测结果应用到实时预警中是下一步将要考虑的工作。

## 参考文献

- [1] LU H, YANG F. Research on Network Traffic Prediction Based on Long Short-Term Memory Neural Network[C]//2018 IEEE 4th International Conference on Computer and Communications (ICCC). Chengdu, China: IEEE, 2018: 1109-1113.
- [2] CHEN G J, LIANG P, WANG K. The Research of Network

Traffic Prediction Model [J]. Information & Communications, 2017, 176(8): 191-194.

- [3] XU S, ZENG B. Network Traffic Prediction Model Based on Auto-regressive Moving Average [J]. Journal of Networks, 2014, 9(3): 97-102.
- [4] SHENG H, ZHANG Y X. Network Traffic Modeling and Forecasting based on ARIMA [J]. Communication technology, 2019, 52(4): 903-907.
- [5] TIAN H, ZHOU X, LIU J. A Hybrid Network Traffic Prediction Model Based on Optimized Neural Network[C]//2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT). IEEE Computer Society, 2017: 284-287.
- [6] LU H P, YANG F. A Network Traffic Prediction Model Based on Wavelet Transformation and LSTM Network [C]//2018 IEEE 9th International Conference on Software Engineering and Service Science. 2018: 1-4.
- [7] BAI X Y, YE X M, JIANG H. Network Traffic Predicting Based on Wavelet Transform and Autoregressive Model [J]. Computer Science, 2007(7): 47-49, 54.
- [8] ZHU Q Y, QIN X Z, JIA Z H, et al. Network traffic prediction based on EMD and particle swarm optimization of LS-SVM [J]. Computer Engineering and Design, 2013, 34(12): 4104-4108.
- [9] MA J Y, WANG P, XIA W, et al. Research on Network Traffic Prediction and Early Warning in Complex Networks [J]. Computer and Modernization, 2018(1): 102-106.
- [10] GAO B, ZHANG Q Y, LIANG Y S, et al. Predicting self-similar networking traffic based on EMD and ARMA [J]. Journal On Communications, 2011, 32(4): 47-56.
- [11] DING X F, ZHAO S H, LI R X, et al. Traffic prediction algorithm of space information network based on combination model [J]. Optical Communication Technology, 2017, 41(7): 44-47.
- [12] LI T, ZHAO C. Nearest Neighbor Optimization k-means Clustering Algorithm [J]. Computer Science, 2019, 46: 216-219.
- [13] LI X G, WEI N, WEI X. A new method for determining parameters of system complexity measures and its application [J]. Systems Engineering-Theory & Practice, 2018, 38(1): 252-262.
- [14] XU K, LI Z Z, LIU L, et al. Network Traffic Prediction based on ARIMA Model [J]. Microelectronics & Computer, 2004, 21(7): 84-87.
- [15] YU F, CHEN D, TANG X. Time Delay Prediction Method Based on EMD and Elman Neural Network [C]//2014 Sixth International Conference on Intelligent Human-Machine Systems and Cybernetics. Hangzhou, China: IEEE, 2014: 368-371.



**YAO Li-shuang**, born in 1993, postgraduate. Her main research interests include machine learning and network management.