

基于双向 GRU 模型的网络流量预测的研究

徐海兵, 郭久明

(迈普通信技术股份有限公司 科技创新部, 四川 成都 610094)

摘要: 当前使用门控循环单元(Gated Recurrent Units, GRU)神经网络进行流量预测时, 普遍存在滞后性以及预测准确性不高的问题, 因此提出一种改进的 GRU 模型进行流量预测的方法。首先基于 GRU 神经网络提出一种双向 GRU 神经网络和人工神经网络堆叠的网络模型, 适用于流量特征、时间特征、事件特征等多维向量的输入; 同时为解决部分时间段准确度不高的问题, 将训练样本进行日期分类, 针对每一类日期生成单独的网络模型, 能大幅提升预测的准确度以及改善预测的滞后性。最后, 为了提升流量峰值的预测准确度, 采用样本的再平衡手段以及自定义损失函数, 实验结果表明, 能较好地达成预期目标。

关键词: 流量预测; 神经网络; 门控循环单元; 损失函数

中图分类号: TN919.2; TP181

文献标识码: A

DOI: 10.16157/j.issn.0258-7998.211517

中文引用格式: 徐海兵, 郭久明. 基于双向 GRU 模型的网络流量预测的研究[J]. 电子技术应用, 2022, 48(2): 19-22, 27.

英文引用格式: Xu Haibing, Guo Jiuming. Research on network traffic prediction based on Bi-GRU model[J]. Application of Electronic Technique, 2022, 48(2): 19-22, 27.

Research on network traffic prediction based on Bi-GRU model

Xu Haibing, Guo Jiuming

(Technological Innovation Department, Maipu Communication Technology Co., Ltd., Chengdu 610094, China)

Abstract: At present, there are some problems such as lag and low prediction accuracy when using gated recurrent units (GRU) neural network to predict traffic. This paper proposes an improved GRU model for traffic prediction. Firstly, based on GRU neural network, a network model integrating Bi-GRU neural network and artificial neural network is proposed, which satisfies the input of multi-dimensional vectors such as traffic features, time features and event features. At the same time, in order to improve the accuracy of some time periods, the training samples are classified into date classes, and a separate network model is generated for each type of date. It can greatly improve the accuracy of prediction and improve the lag of prediction. Finally, in order to improve the prediction accuracy of peak traffic, the experimental results show that the proposed goal can be achieved by the means of sample propensity balance and user-defined loss function.

Key words: traffic prediction; neural network; gated recurrent unit; loss function

0 引言

随着网络的普及, 网络流量的规模不断被刷新, 高效且合理地利用网络资源变得尤为重要。一方面, 网络资源分配的不合理可能导致部分网络资源由于耗尽而无法正常使用, 甚至造成网络瘫痪, 而其他链路资源可能却处于过剩的状态, 严重影响了用户的上网体验; 另一方面, 虽然在前期合理分配了网络资源, 但网络流量具有突发性, 原本充足的网络资源可能出现短缺的情况。为了解决此问题, 现有软件定义网络(Software Defined Network, SDN)^[1]控制器会实时检查链路状况, 在一定程度上缓解了网络拥塞, 但由于调度时已经发生了拥塞, 无法满足更高等级、更好服务质量的要求。鉴于此, 如果能够精准预测网络流量, 提前发现未来时刻的网络流量变化情况, 流量调度系统则可以提前进行合理调度, 有

效避免拥塞的发生。

1 相关工作

1.1 基于统计的模型

基于统计的模型通常先采用差分法对网络流量原始数据平稳化处理, 提取网络流量数据的自相似特征, 然后将平稳后的数据通过统计模型进行拟合和检验^[1], 例如 ARMA、ARIMA^[2]等统计模型。统计模型本身十分简单, 只需要内生变量而不需要借助其他外生变量, 但统计模型只能捕捉线性关系, 而不能捕捉非线性关系; 同时要求自身或差分之后的数据是平稳的, 例如无法用统计模型预测股票类型的数据, 因为数据是非平稳的且受政策和新闻等其他事件影响^[3]。

1.2 基于机器学习的模型

基于机器学习的流量预测模型包括梯度提升决策

树^[4]、回归模型^[5]以及 boosting 模型^[6]等,首先需要对其原始的数据进行分类、统计、补缺、寻找等操作,得到时间差异序列、时间趋势序列和时间序列,以此增加样本特征的多样性^[6]。虽然机器学习模型对线性或非线性的规律有较强的拟合能力,但需要挖掘尽可能多的样本特征,增加了模型开发的难度。

1.3 基于经验模态分解的模型

为了改善流量预测的滞后性,一些文献提出经验模态分解(Empirical Mode Decomposition, EMD)^[7]和集成经验模态分解(Ensemble Empirical Mode Decomposition, EEMD)^[8]的理论方法,其方法为将访问流量数据分成若干个 IMF 分量和残余分量 Res,之后针对 IMF 分量和 Res 分量分别运用统计模型、机器模型或神经网络模型进行预测,最后将各自的预测结果进行叠加重构^[9]。本文针对该方法进行实验,结果表明该理论方法虽然对规律性较强的流量数据有较好的滞后性改善,但对变化幅度较大的流量数据存在 IMF 分量数量不确定的问题,如果强行指定流量数据分解的 IMF 数量,则存在流量预测的准确性不高的问题。

1.4 基于循环神经网络的模型

由于循环神经网络 RNN 考虑了时间序列之前的信息,故非常擅长处理序列数据,已在众多领域(包括语音识别、文档识别、手写识别及图像分析)中得到广泛应用^[10]。长短期记忆循环神经网络(Long Short-Term Memory, LSTM)在解决了梯度消失和爆炸问题后,能够更好地进行分类和预测任务^[11]。门控循环单元 GRU 是 LSTM 的一种变体,保持了 LSTM 效果的同时又使结构更加简单,能够很大程度提高训练的速率^[12]。

2 流量预测模型

2.1 基于循环神经网络 GRU 预测过程

首先对某公司的互联网出口的网络流量进行聚合形成流量时间序列,对流量时间序列的异常值采用孤立森林算法去除,对流量时间序列的缺失值采用线性插值法补齐;然后分别对最近一天的流量时间序列、最近第二天的流量时间序列、最近一周前相同时间的流量时间序列分别滑动 48 个时间窗口,最终形成维度为 48×3 的样本特征 $[[X_t, X_{t-48}, X_{t-336}], \dots, [X_{t-1}, X_{t-49}, X_{t-337}],$

$[X_t, X_{t-48}, X_{t-336}]$, 样本标签 Y_t 的值等于 X_{t+1} 。将样本特征和样本标签形成的样本集划分为训练集和测试集,采用 K 折交叉验证,在精确性方面采用均方根误差(Root Mean Square Error, RMSE)和平均百分比误差(Mean Absolute Percentage Error, MAPE)作为评估函数。

$$MAPE = \sum_{t=1}^n \frac{(\text{Predicted}^t - \text{observed}^t)}{\text{observed}^t} \frac{100}{n} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (\text{Predicted}^t - \text{observed}^t)^2}{n}} \quad (2)$$

$$DIFF = \frac{\sum_{t=1}^n (\text{Predicted}^{t+1} - \text{observed}^t)}{n} \quad (3)$$

其中, n 为测试集的个数; Predicted^{t+1} 为 $t+1$ 时刻的预测值, observed^t 为 t 时刻的真实值;自定义差异性的度量函数 DIFF 用来度量 $t+1$ 时刻预测值与 t 时刻真实值之间的差异性,如果测试集的波动性较强且 DIFF 较小,则可能存在滞后性。需要注意的是,针对波动较小的流量曲线,则 DIFF 的数值无法反映滞后性。

2.2 预测结果以及分析

测试集 2020.9.29 日 20:30:00 到 2020.10.01 日 00:00:00 之间预测值和真实值如图 1 所示,横轴代表为流量发生的时刻,其上标代表星期,如 21² 代表星期二的 21:00 时;纵轴代表为流量值的大小,预测结果 MAPE=27.55, MSE=238.15, DIFF=43。从准确性方面来看,预测值与真实值的趋势较为一致,但从细节中不难发现流量极值部分预测存在明显的滞后性,即流量的 $t+1$ 的预测结果和 t 时刻的真实值较为一致。

3 改进的双向 GRU 模型的流量预测方法

从上文可以看出,基于 GRU 的流量预测模型虽然在准确度方面基本满足要求,但预测时存在较为严重的滞后性问题,而流量预测的滞后性无论是对智能网络的故障预测还是针对 SDN 的流量调度均存在致命影响,流量预测的滞后性会导致流量预测变得毫无意义。鉴于此,流量预测的滞后性成为流量预测过程中的必须解决的一个难题。本文重点从多维特征输入、基于日期的模型分类、自定义损失函数等方面来改善流量预测的滞后

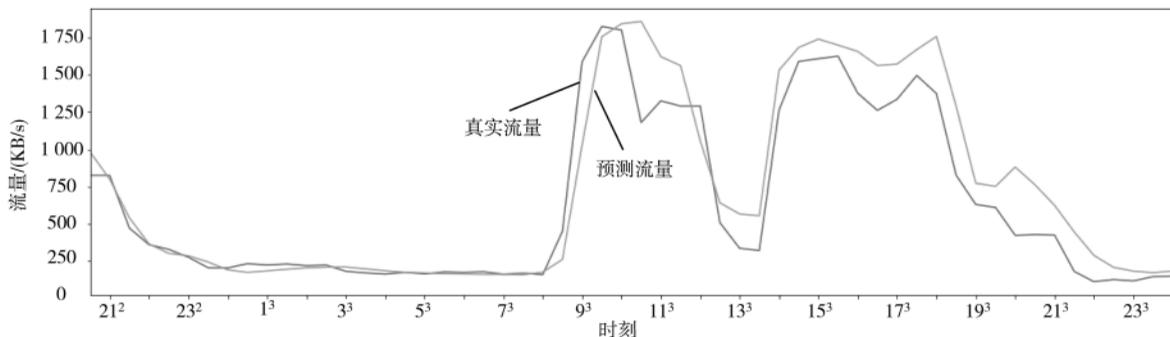


图 1 GRU 模型测试结果

性问题,同时提高流量预测的准确度。

3.1 多维特征输入

基于 GRU 的流量预测模型存在滞后性问题主要是由于流量的自身规律性不明显,模型倾向性加大了上一个时间步的流量大小所对应的参数权重。通过对流量的规律性进行分析发现,流量呈现每周和每天的规律性,如图 2 所示,横坐标代表的是时刻,其上标代表星期,坐标点间隔 3 小时,其中周一到周五 8:30~18:00 是工作时间,其他是非工作时间。

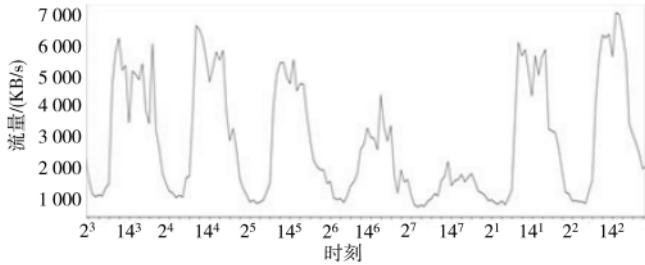


图 2 流量的规律性

根据上述分析,增加流量的时间属性作为样本特征,具体为将样本标签对应的时间分解为星期、小时、分钟,同时针对流量存在突发的规律,增加流量的事件特征,根据流量的突发性大小共分为高、中、低 3 种事件类型,例如是否有视频会议、是否有集体性活动等,将时间特征和事件特征进行独热编码,输入全连接神经网络^[13],如图 3 所示。经过多维特征的模型输入,降低上一个时间步的参数权重,从而改善流量预测的滞后性。

将 2020 年 8 月 29 日到 2020 年 9 月 30 日之间的流量作为样本集,其中前几个星期的流量作为训练集,最后一个星期的流量作为测试集,将每个样本的 144 维的特征向量输入到双向循环网络,将每个样本的 33 维的时间特征向量以及 3 维的事件特征向量输入到全连接神经网络。经过测试发现滞后性明显改善,与此同时准确性也有所提高,输出结果为 MAPE=19.16,RSME=200.76,DIFF=28.59,如图 4 所示。

3.2 基于日期的模型分类

考虑到非工作日的权重参数可能影响到工作日的

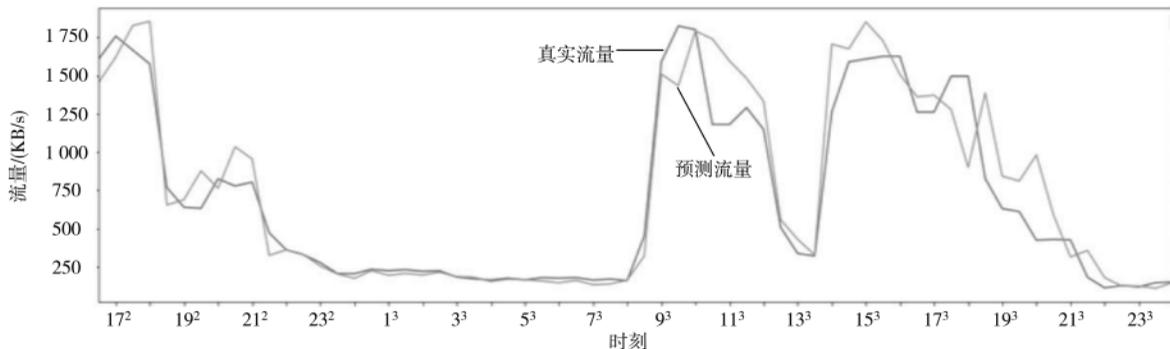


图 4 多维特征测试效果图(局部)

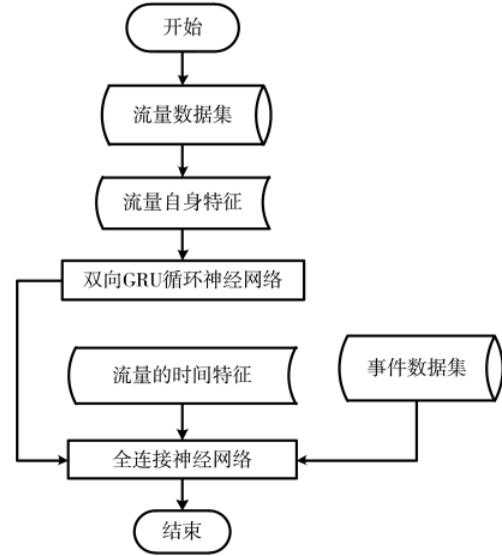


图 3 多维特征流程图

模型参数,本文将流量数据集的时间是否为工作日进行分类,根据分类结果分别建立工作日模型和非工作日模型,如图 5 所示。

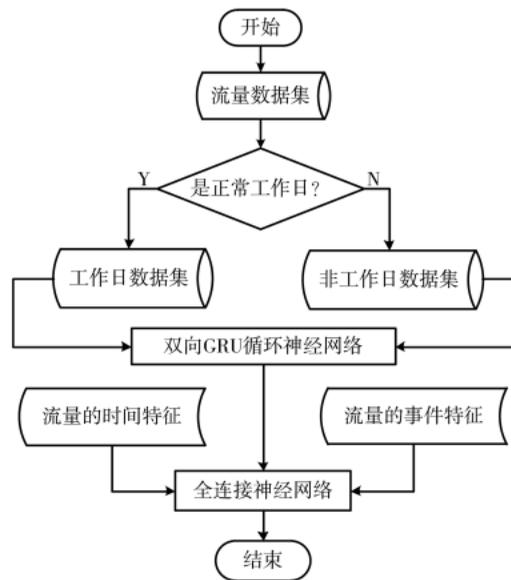


图 5 日期模型的流程图

工作日模型的输入是 31 维的时间特征向量,非工作日模型输入是 29 维的时间特征向量,法定节假日归属于非工作日类型,流量特征向量和特征向量同 3.1 小节所述。测试验证的结果为 MAPE=12.2, RMSE=100.63, DIFF=34.4, 从图 6 以及统计数据不难看出, DIFF 增大而 MAPE 值和 RMSE 值降低,说明整体上滞后性问题有所改善且准确性进一步提升。

3.3 自定义损失函数

在流量预测的过程中,尤其是在 SDN 的调度场景中,对流量的峰值预测结果较为关注,预测值应稍大于实际值为最佳。然而现有的样本存在不平衡现象,即大部分样本分布在非工作时间对应的低值部分,导致峰值预测精度不高。为了解决此问题,一方面从样本不平衡的角度出发,增强极大值的样本数量,具体为通过重采样极大值范围的样本特征,然后将样本标签加入一定的白噪声,新的样本标签 $Y_{new}=Y_{old}+N_t$, 其中 Y_{old} 是重采样前的样本标签, Y_{new} 是采样后的样本标签, N_t 取 $[0, Y_{old} \times 10\%]$ 区间内的白噪声。为了进一步加大极大值预测错误的惩罚,自定义模型的损失函数:

$$LOSS(x, f(x)) = \begin{cases} \sum_{t=0}^n \frac{1}{2} (y_t - f(x))^2 & y \geq \frac{\lambda \cdot \sum_{t=0}^n y_t}{n} \\ \sum_{t=0}^n |y_t - f(x)| & y < \frac{\lambda \cdot \sum_{t=0}^n y_t}{n} \end{cases} \quad (4)$$

其中, y_t 为样本集的标签值, $f(x)$ 为模型的预测值; λ 用来调节极大值的范围, 本文选择 $\lambda=1.2$ 。使用与 3.1 小节同样的数据集, 采用上述的样本不平衡方法以及自定义损失函数进行模型训练和预测, 测试结果为 MAPE=15.93, RMSE=152.16。从图 7 中不难看出, 峰值预测的准确性进一步提高, 但极小值部分的预测准确性有一定的下降, 整体上符合预期。

4 结论

本文选择循环神经网络 GRU 作为流量预测的理论模型, 针对预测过程普遍存在的滞后性以及准确性不高的问题, 提出一种改进后的双向循环神经网络 GRU 模

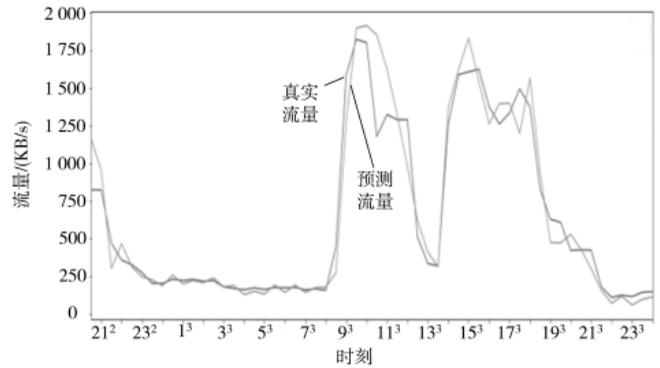


图 7 自定义损失函数的测试效果图

型, 通过对预测过程中的结果分析, 针对性采取多维特征输入、基于日期的模型分类、自定义损失函数多个步骤改进预测的结果, 其中多维特征输入和基于日期的模型分类对流量预测的滞后性有明显改善, 能大幅提升预测的准确性, 样本再平衡以及自定义损失函数对流量峰值的预测能够达到预期的目标。后续的研究中将探索使用注意力机制^[14]以及基于时间的卷积神经网络 TCN^[15]对流量预测的精度进一步进行提升。

参考文献

- [1] XIE X, WU J. Real-time flow identification based on neural network and OpenFlow over SDN[C]//2018 10th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2018.
- [2] 张冉, 赵成龙. ARIMA 模型在网络流量预测中的应用研究[J]. 计算机仿真, 2011, 28(2): 171-174.
- [3] 李士宁, 闫焱, 覃征. 基于 ARIMA 模型的网络流量预测[J]. 计算机工程与应用, 2006(29): 148-150.
- [4] 康传利, 顾峻峰, 刘兆威. 梯度提升回归树的旅游流量预测模型[J]. 数学的实践与认识, 2019, 49(15): 251-261.
- [5] VLAHO E, KARL G, GOLIAS J. Short-term traffic forecasting: where we are and where we're going[J]. Transportation Research Part C: Emerging Technologies, 2014, 43: 3-19.
- [6] 童行行. 基于机器学习的网络流量分析研究[D]. 北京: 清华大学, 2015.

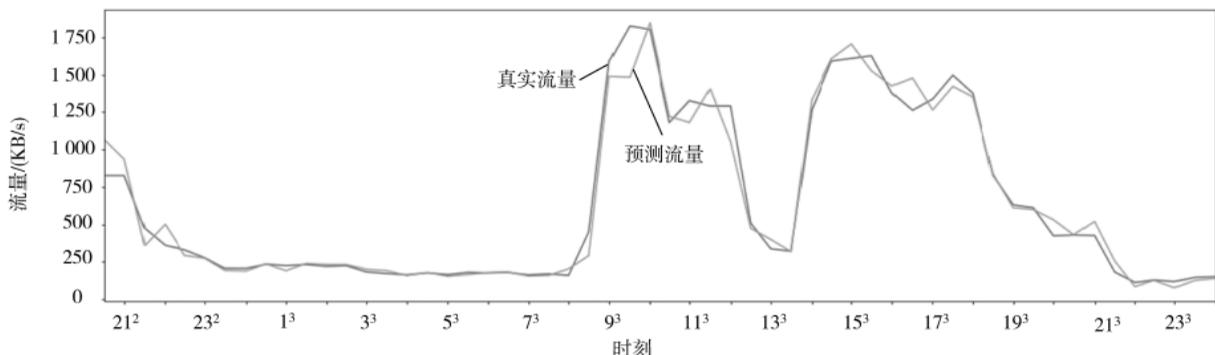


图 6 日期模型的测试效果图

(下转第 27 页)

4 结论

本文研究设计了一种新型自毁监测和自毁执行系统电路。新型自毁监测电路包括封装拆卸监测电路和上电时序监测电路,可以对被保护芯片进行实时保护监测,电路简易,功耗低。新型自毁执行电路利用超级电容和高能膜构成,在断电的情况下,自毁执行系统依然可以执行自毁,保护芯片内部存储的信息。自毁监测和执行系统的自毁延时可以小于 1 ms,实现了实时监测、快速自毁的低功耗自毁监测和执行。

参考文献

- [1] 王帅,高秀峰,李玺,等.涉密 U 盘自毁装置:CN 204657084 U[P]. 2015-09-23.
- [2] CHRISTOPHER K. This message will self-destruct: the growing role of obscurity and self-destructing data in digital communication[J]. Bulletin of the American Society for Information Science and Technology(Online), 2013, 40(2).
- [3] 赵玉清,孙良成,李建强,等.集束子弹药引信独立自毁自失效设计方法[J].兵器装备工程学报, 2018, 39(12): 41-45.
- [4] KIM Y K, HONG C S. Threshold estimation in self-destructing scheme using regression analysis[C]//2017 19th Asia-Pacific Network Operations and Management Symposium(APNOMS), 2017: 135-138.
- [5] Self-destructing silicon chip[J]. Materials Today, 2002, 5(3).
- [6] 姜冬,王慧强,冯光升,等.基于模糊层次化评估的分布式系统自毁感知方法及应用[J].小型微型计算机系统, 2012(4).
- [7] CAO Q, CHENG K, LI Z, et al. Prevention of reverse engi-

neering of security chips: U.S. Patent 9,853,001[P]. 2017-12-26.

- [8] FU X, WANG Z, WU H, et al. How to send a self-destructing email: a method of self-destructing email system[C]//2014 IEEE International Congress on Big Data, 2014: 304-309.
- [9] 衡立业. 数据加密和异常数据自毁技术在网络信息安全中的研究[J]. 网络安全技术与应用, 2020(6): 35-36.
- [10] GU X, LOU W, SONG R, et al. Simulation research on a novel micro-fluidic self-destruct device for microchips[C]//2010 IEEE 5th International Conference on Nano/Micro Engineered and Molecular Systems, 2010: 375-378.
- [11] JEFF A, STEVEN T, JEFFREY J. Self-destructing electronic device: U.S. Patent 9812407[P]. 2017-11-07.
- [12] ZENG L, CHEN S, WEI Q, et al. Se Das: a self-destructing data system based on active storage framework[C]//2012 Digest APMRC, 2012: 1-8.

(收稿日期: 2021-05-31)

作者简介:

胡长征(1996-),男,硕士研究生,主要研究方向:数字集成电路、嵌入式系统等。

马伟(1995-),男,博士研究生,主要研究方向:模拟射频集成电路设计。

胡伟波(1982-),通信作者,男,博士,研究员,主要研究方向:模/数和数/模转换器、无线感知芯片和系统实现、第三代半导体、生物芯片, E-mail: weibohu@hotmail.com。



扫码下载电子文档

(上接第 22 页)

- [7] 赵雪花,陈旭袁,旭琦.基于 EMD 的数据驱动模型在径流预测中的应用[J].系统工程, 2014, 32(9): 150-154.
- [8] 司友强,吴润华,施鹏程.基于 EMD, EEMD 与 CEEMD 的信号时频分析技术对比研究[J].CT 理论与应用研究, 2019, 28(4): 417-426.
- [9] WU Y, WU Q, ZHU J. Improved EEMD-based crude oil price forecasting using LSTM networks[J]. Physical A: Statistical Mechanics and Its Applications, 2019, 516: 114-124.
- [10] 李洁,林永峰.基于多时间尺度 RNN 的时序数据预测[J].计算机应用与软件, 2018, 35(7): 33-37.
- [11] YU Y, SI X, HU C, et al. A review of recurrent neural networks: LSTM cells and network architectures[J]. Neural Computation, 2019, 31(1): 1-36.
- [12] 超人汪小建. GRU 神经网络[EB/OL]. (2017-08-17)[2021-

03-16]. https://blog.csdn.net/wangyangzhizhou/article/details/77332582.

- [13] 陈雯柏,吴细宝,陈启丽.人工神经网络原理与实践[M].西安:西安电子科技大学出版社, 2016: 44-47.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. ArXiv: 1706.03762, 2017.
- [15] BAI S, KOLTER J Z, KOLTUN V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. ArXiv: 21803.01271, 2018.

(收稿日期: 2021-03-16)

作者简介:

徐海兵(1982-),男,硕士,高级工程师,主要研究方向:时序预测、智能运维、OCR、视频理解。

郭久明(1982-),男,本科,工程师,主要研究方向:数据通信网络、意图驱动网络。



扫码下载电子文档

版权声明

经作者授权，本论文版权和信息网络传播权归属于《电子技术应用》杂志，凡未经本刊书面同意任何机构、组织和个人不得擅自复印、汇编、翻译和进行信息网络传播。未经本刊书面同意，禁止一切互联网论文资源平台非法上传、收录本论文。

截至目前，本论文已经授权被中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普网）、DOAJ、美国《乌利希期刊指南》、JST 日本科技技术振兴机构数据库等数据库全文收录。

对于违反上述禁止行为并违法使用本论文的机构、组织和个人，本刊将采取一切必要法律行动来维护正当权益。

特此声明！

《电子技术应用》编辑部

中国电子信息产业集团有限公司第六研究所