

基于 BPNN-LSTM 组合模型的云计算资源负荷预测

陈冬林, 林文健, 黄莉莉

(武汉理工大学 湖北省电子商务大数据工程技术中心, 湖北 武汉 430070)

摘 要:有效的负荷预测对提高云计算资源利用效率、降低电力消耗和运维人力资源成本具有重要意义。针对当前云计算资源负荷预测精度低、难以准确把握历史数据时序特征的问题,提出一种引入时序因素的 BPNN-LSTM(BP 神经网络-长短期记忆网络)组合预测模型。综合考虑对云计算资源负荷具有重要影响的日期和时间因素,采用 LSTM 对 BPNN 的预测残差进行修正,对不同时间维度的云计算资源负荷进行预测。通过与多种预测模型的对比实验,验证了所提出的模型在对云计算资源负荷预测上具有更高的精度和稳定性。

关键词:云计算资源;负荷预测;时序因素;BP 神经网络;长短期记忆网络

中图分类号: TP3

文献标识码: A

文章编号: 1008-7729(2020)01-0053-08

一、引 言

云计算对于计算资源的公共化和开放化具有重要意义,但偏低的利用率一直是云计算资源管理方面的难题。研究显示,目前,各类云计算资源利用率不足 50%,大量计算资源处于闲置状态^[1]。而维持较多云计算资源的正常运转一方面需要消耗大量电力资源,另一方面基础设施的管理和维护也会产生较高的人力成本。云计算资源偏低的利用率已然造成了社会资源的巨大浪费。

提升云计算资源利用率的有效方法是对资源的负荷量进行精准预测,通过解析云计算资源使用情况的历史数据,探寻负荷量变化的内在规律,预测未来一段时间内云计算资源的负荷量。云服务供应商可以根据预测结果动态调整资源配置,为用户提供不同服务等级(Service-Level Agreement, SLA)的云计算资源,降低自身电力资源和人力资源成本。

鉴于云计算资源利用率偏低的现状及其负荷预测能够带来巨大收益,云计算资源负荷的准确预测问题已经引起学术工作者和云计算行业的广泛关注^[2]。近年来,国内外学者对如何有效提高云计算资源负荷预测的精度和稳定性进行了探索和研究,研究方法主要分为序列分析类和机器学习类。云计算资源负荷预测的序列分析方法有:自回归模型^[3]、Holt-Winters 模型^[4]、经验模态分解(EMD)^[5]等。随着对云计算资源负荷预测要求的不断提高,序列分析类方法在预测精度和复杂环境下的预测稳定性问题逐渐凸显。机器学习的兴起使云计算资源的负荷预测进入新的领域,采取的主要方法有神经网络模型、支持向量机方法、粒子群优化算法等。Kumar 等^[6-7]基于 BP 神经网络对云数据中心的工作负载进行预测;Nhuan 等^[8-9]基于长短期记忆神经网络构建云数据中心工作量预测模型,对历史负荷数据的时间序列进行分析,实现预测的高精确度;孟煜等^[10]提出一种云服务用户并发量区间预测模型,利用梯度下降粒子群优化的支持向量机作为主要预测方法。

现有云计算资源负荷预测的研究中,缺乏对时序因素的考虑,未能基于当前的时间节点对云计算资源负荷进行考量,难以把握云计算资源负荷的动态时序特征。基于上述分析,笔者提出一种引入时

收稿日期:2019-07-23

基金项目:国家重点研发计划(2018YFB1404303),湖北省自然科学基金资助项目(20181j0034)

作者简介:陈冬林(1970—),男,湖北安陆人,教授,博士生导师

序因素的 BP 神经网络与长短期记忆网络相结合的 BPNN-LSTM 组合模型,对云计算资源负荷进行预测,以期在不同时间维度上获得更好的预测精度。

二、理论基础

本文以 BP 神经网络(Back Propagation Neural Network, BPNN)和长短期记忆网络(Long Short Term Memory networks, LSTM)为理论基础,对不同时间维度的云计算资源负荷进行预测。

(一) BP 神经网络模型

BP 神经网络^[11]是一种信号正向传递、误差逆向传播的多层前馈神经网络,广泛应用于函数逼近、模式识别等领域。BP 神经网络的拓扑结构包括输入层、隐含层和输出层,每层均由神经元联结组成。BP 神经网络的正向计算过程中,信息由输入层输入,经各隐含层的处理和计算,由输出层输出处理结果。误差的逆向传播中,计算输出层处理结果的误差,反向传播误差信号,采用梯度下降的方式修正各神经元间联结的权值,实现网络优化。

(二) LSTM 网络模型

LSTM 网络^[12]是一类可以处理数据长期依赖问题的特殊神经网络,克服了 BP 神经网络求解过程中的梯度消失或梯度爆炸问题。LSTM 的隐含层不再是普通的神经元,而是用于增加或删除细胞状态(cell state)中信息的三重门结构,其网络结构如图 1 所示。

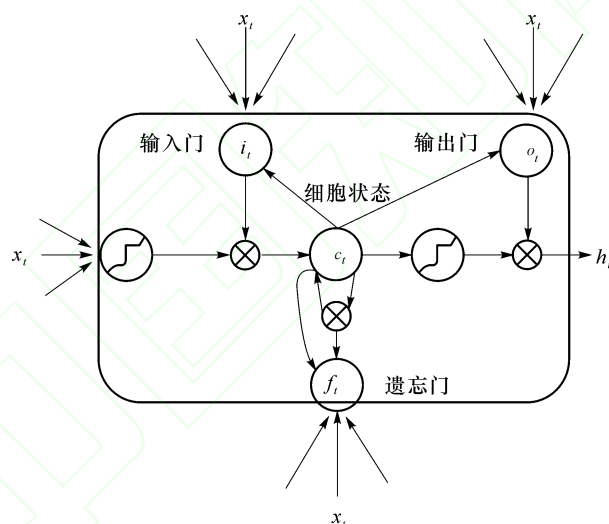


图 1 LSTM 网络结构

LSTM 的前向计算过程中,每一个记忆细胞(cell)的运算均由输入门(input gate)、遗忘门(forget gate)、输出门(output gate)控制,每个门依据激活函数的判断逻辑决定是否将运算结果向下传递。当前状态下,输入门 i_t 控制输入数据 x_t 新生成的信息是否可以加入到当前细胞状态 c_t ;遗忘门 f_t 控制上一时刻细胞状态 c_{t-1} 的信息是否可以加入到当前细胞状态;输出门 o_t 基于输入门和遗忘门更新后的细胞状态,输出隐藏状态 h_t 。LSTM 误差的反向误差传播程与 BP 神经网络类似,通过计算误差,按照时间和网络层级两个方向反向传播误差信号,更新权值梯度,优化联结权值。

三、引入时序因素的 BPNN-LSTM 组合预测模型

针对引言中提出的问题,提出引入时序因素的 BPNN-LSTM 组合预测模型,具体框架如图 2 所示。首先,基于云计算负荷历史数据匹配相应的时序数据,经过对数据的清洗和数据归一化处理,得到增

强数据集，作为组合模型的输入数据；再拆分输入数据，形成 BPNN 模型的训练数据和测试数据，确定 BPNN 网络结构和权值、阈值参数，训练网络，对测试数据进行预测，计算预测误差；然后，基于 BPNN 误差，再次匹配时序数据，并进行归一化处理，作为 LSTM 网络的输入数据，确定 LSTM 网络结构和权值、阈值参数，进行训练和预测，得到 BPNN 残差预测值；最后，加总 BPNN 预测结果和 LSTM 残差预测结果，得到最终的云计算资源负荷预测值。

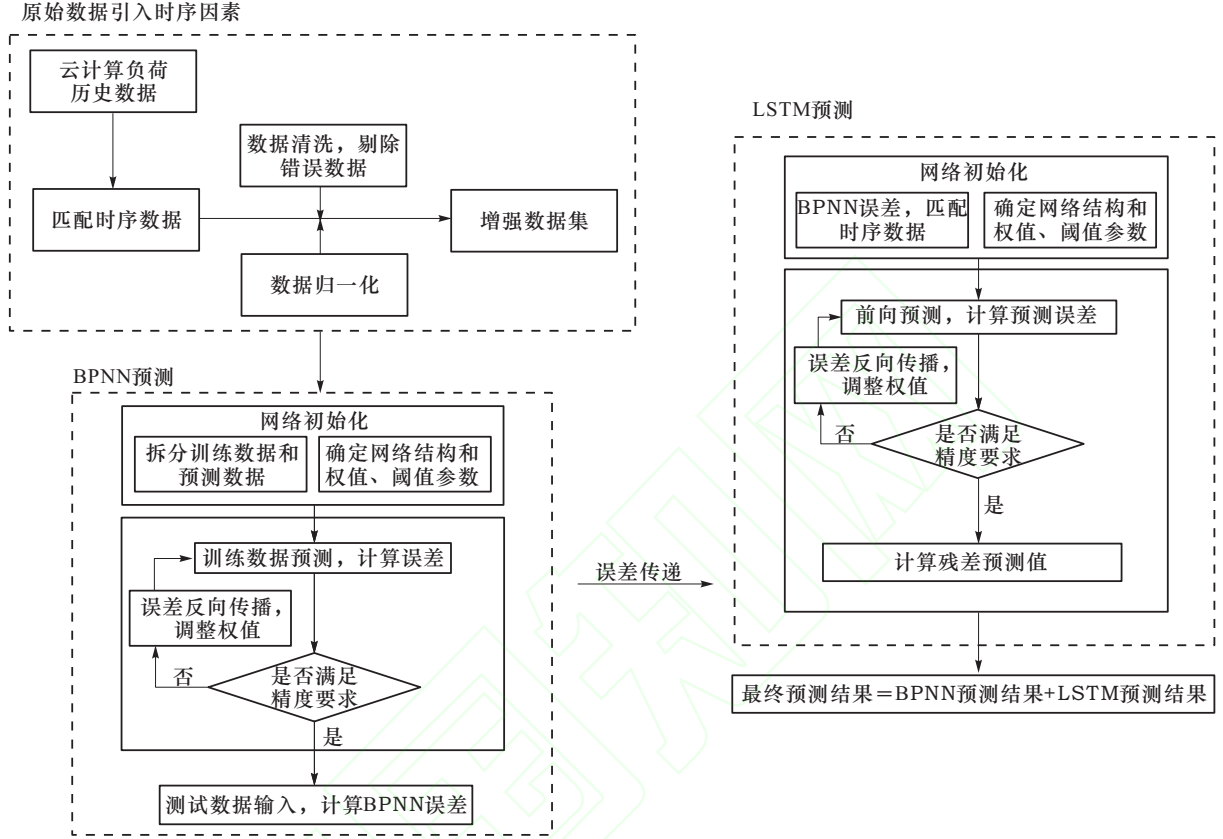


图2 引入时序因素的 BPNN-LSTM 组合模型预测流程

(一) 数据输入及其处理

以 $x_i (i \in (1, n+k))$ 表示第 i 个时间节点的云计算资源负荷历史数据，以 x_1, x_2, \dots, x_n 构建第 1 个时间节点到第 n 个时间节点的历史数据，以 x_2, x_3, \dots, x_{n+1} 构建第 2 个时间节点到第 $n+1$ 个时间节点的历史数据，依此类推，构建云计算资源负荷历史数据矩阵

$$\mathbf{X}_{\text{Original}} = \begin{pmatrix} x_1 & \cdots & x_k \\ \vdots & & \vdots \\ x_n & \cdots & x_{n+k-1} \end{pmatrix} \quad (1)$$

基于企业用户云服务需求的时序特征，引入时序因素的时间属性 D_i (单位: d, $i \in (1, n+k)$) 和时间属性 T_i (单位: h, $i \in (1, n+k)$)，匹配云计算资源负荷历史数据矩阵，构建数据集 \mathbf{X} 和待预测数据集 \mathbf{Y} 。

$$\mathbf{X} = \begin{pmatrix} x_1 & \cdots & x_k & D_1 & T_1 \\ \vdots & & \vdots & \vdots & \vdots \\ x_n & \cdots & x_{n+k-1} & D_n & T_n \end{pmatrix}, \quad \mathbf{Y} = \begin{pmatrix} x_{k+1} \\ \vdots \\ x_{n+k} \end{pmatrix} \quad (2)$$

经数据清洗和归一化处理后，得到增强数据集 (\mathbf{X}, \mathbf{Y}) ，作为组合模型的输入数据。

(二) BP 神经网络预测

拆分输入数据 (\mathbf{X}, \mathbf{Y}) 为训练数据集 $(\mathbf{X}_{\text{train}}, \mathbf{Y}_{\text{train}})$ 和测试数据集 $(\mathbf{X}_{\text{test}}, \mathbf{Y}_{\text{test}})$ ，构建输入层、第一层

隐含层、第二层隐含层、输出层的四层网络结构,采用随机函数生成的方式初始化权值和阈值参数。选取 sigmoid 函数为 BPNN 神经网络的激活函数,函数表达式为

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

对训练数据集进行正向预测,设置比对预测结果 Y_{predict} 和实际负荷数据 Y_{test} 的损失函数为均方误差平方根函数

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^m (Y_{\text{train},i} - Y_{\text{predict},i})^2 \quad (4)$$

依据梯度下降规则调整权值,将输出层的反传误差信号经第二层隐含层、第一层隐含层传递计算,最终传至输出层。通过误差的反向传播,修正各个节点之间的连接权值,实现网络优化。

网络训练完成后,代入测试数据集 X_{test} ,计算 BPNN 预测结果 Y_{BP} 与实际负荷数据 Y_{test} 的残差 E ,并将其传向 LSTM 网络。

(三) LSTM 网络预测

基于 BPNN 预测残差 E ,再次匹配时序数据 D_i 和 T_i ,经归一化处理,作为 LSTM 网络的输入数据。输入数据经输入门、遗忘门、输出门的运算,生成当前细胞状态保留信息 o_i 和隐藏状态的输出 h_i ,并将输出数据向下一时刻传递,最终得到残差预测结果 E' 。

最终预测结果 $Y_{\text{BP_LSTM}}$ 为 BPNN 预测结果和 LSTM 残差预测结果之和,即

$$Y_{\text{BP_LSTM}} = Y_{\text{BP}} + E' \quad (5)$$

采用平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)评价模型预测结果,函数表达式为

$$\text{MAPE} = \frac{1}{n} \sum_{t=k+1}^{n+k} \frac{|Y_{\text{BP_LSTM}}(t) - Y(t)|}{Y(t)} \quad (6)$$

四、数值实验

(一) 数据来源及其时序特征分析

本文使用的数据是 Wikipedia 网站 2012 年 1 月 1 日至 2015 年 12 月 31 日每日 24 小时的云计算资源负荷数据^[13](并发量数据),图 3(a)和图 3(b)描述了该网站云计算资源负荷的周数据和日数据。

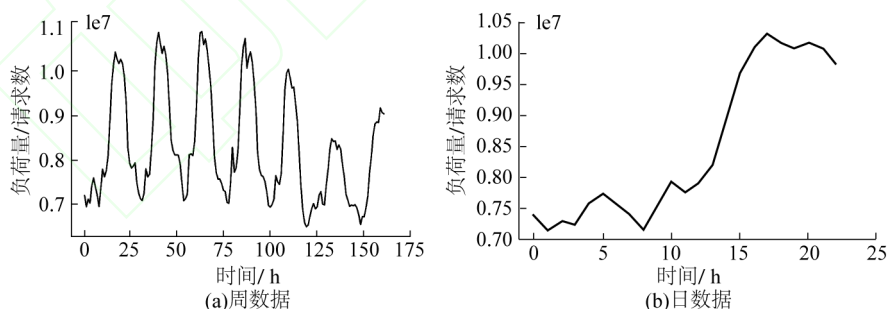


图 3 云计算资源负荷数据

如图 3 所示,云计算资源负荷量在月度上有较为明显的变化规律,周与周之间的变化规律基本一致。在一周的时间维度上,可以看出工作日的负荷量明显高于周末的负荷量,初步判断,日期属性在一定程度上会影响云计算资源负荷。在日变化曲线上可以看出,负荷量高峰出现在每日的约 18:00—22:00,低谷期则出现在每日的约 0:00—7:00,判断时间属性与云计算资源负荷有较强的相关性。

经测算,云计算资源负荷量与日期因素、时间因素的 Pearson 相关性检验均在 0.01 的水平上显著相关。其中,日期因素的 Pearson 相关系数为 0.485,与云计算资源负荷量中等强度相关;时间因素的 Pearson 相关系数为 0.715,与云计算资源负荷量强相关。结合曲线变化趋势分析和 Pearson 相关性检验

结果可知，用户对云计算资源的需求与时序因素相关性较大。

(二) 基于 BPNN-LSTM 组合模型的云计算资源负荷预测

基于 2015 年 1 月 1 日至 11 月 1 日的 7 319 条云计算负荷数据，匹配时序数据，经数据清洗和归一化处理，作为模型的训练集；基于 2015 年月 11 月 2 日至 11 月 3 日的 48 条云计算负荷数据，匹配时序数据，经数据清洗和归一化处理，作为模型的测试集。设置 BPNN 为 4 层网络结构，包含两层隐含层，设置输入层 24 个神经元，输出层 1 个神经元。因隐含层神经元个数对实验结果有较大影响，为提高预测精度，基于文献[2]的方法，对 BPNN 第一层隐含层和第二层隐含层的节点数进行实验，实验结果如图 4 所示。由实验结果可选择使得预测误差最小的神经元个数方案，最终确定 BPNN 第一层隐含层神经元个数为 13 个，第二层隐含层神经元个数为 8 个。

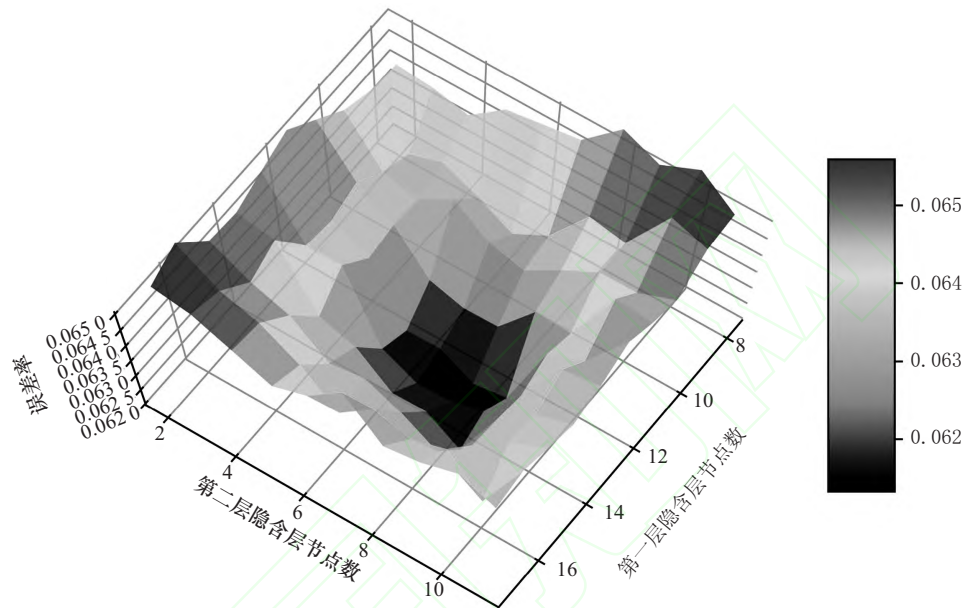


图 4 BP 神经网络隐含层节点数对应的误差率

综合考虑预测精度和预测时间，设置 BPNN 训练步长为 25 000，学习率为 0.095；设置 LSTM 训练规模 batch_size 为 40，迭代次数为 1 000。BPNN-LSTM 组合模型的预测结果如图 5(a) 所示。由图 5 可知，所提出的组合模型具有良好的预测效果，在 48 小时的时间维度上较好地拟合了云计算资源负荷，与真实的负荷变化趋势基本一致。BPNN-LSTM 组合模型预测的平均误差率为 3.524%，误差曲线如图 5(b) 所示。虽然有部分点的误差率超过 10%，但大部分误差基本控制在 -5%~5% 之间，拟合效果较好。

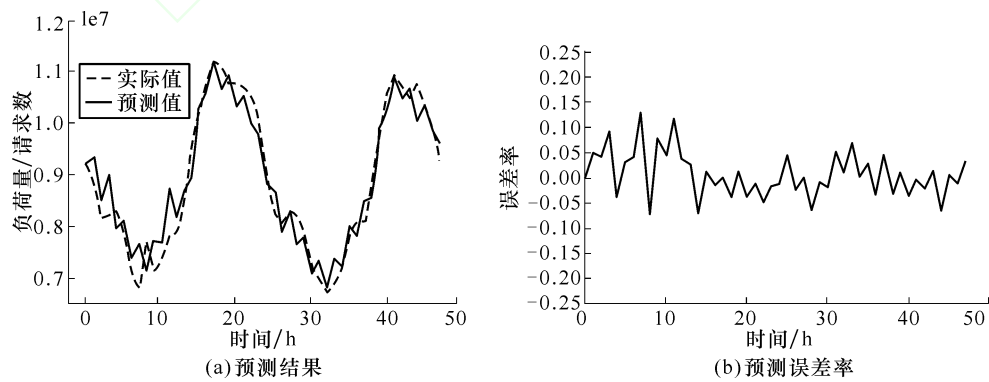


图 5 BPNN-LSTM 组合模型 48 小时预测结果和误差率

为检验所提出的模型在不同时间维度上预测效果的稳定性，分别对未来 24 h、48 h、168 h、720 h 的云计算资源负荷进行预测，预测结果如图 6 所示。模型在预测时间维度为 24 h、48 h、168 h 和 720 h

上预测误差分别为 4.228%、3.524%、4.735% 和 3.978%，总体预测精度较高，且具有较好的稳定性。

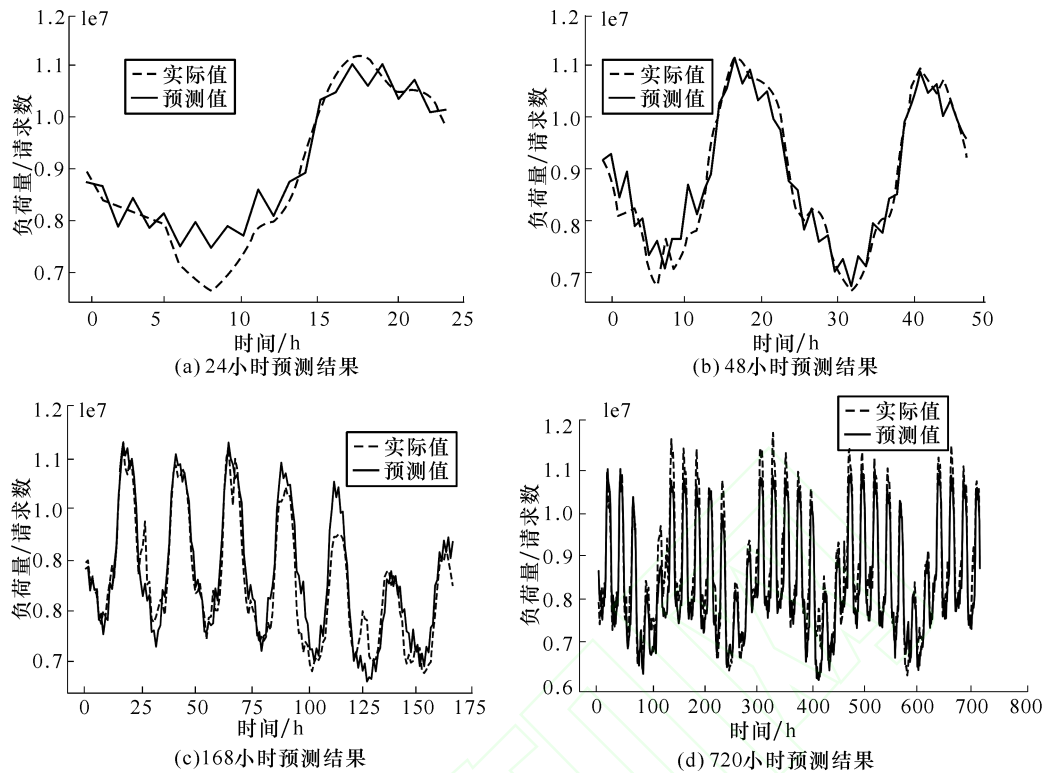


图 6 BPNN-LSTM 组合模型预测结果

(三) 实验结果分析与对比

为综合考察所提出的 BPNN-LSTM 预测模型和其他预测模型的预测精度及在不同时间维度上预测效果的稳定性，基于相同数据集，采用本文所提出的引入时序因素的 BPNN-LSTM 模型、普通的 BPNN-LSTM 模型、GA-BP 模型、BPNN 模型、LSTM 模型对未来 24 h、48 h、168 h、336 h 和 720 h 的云计算资源负荷进行预测。表 1 记录了不同模型在不同时间维度上的误差率数据。

表 1 各模型不同时间维度上预测的误差率

%

时间维度/h	BP-LSTM(引入时序)	BP-LSTM	GA-BP	BPNN	LSTM
24	4.23	6.93	7.52	9.74	9.40
48	3.52	8.42	9.42	10.25	10.37
168	4.74	8.63	8.13	8.58	8.75
336	4.37	6.27	8.57	11.36	9.48
720	3.98	7.84	7.79	9.73	9.15

实验结果显示：(1)引入时序因素的 BPNN-LSTM 模型在预测误差上明显低于其他模型，且在不同时间维度上预测误差也均维持在 4.8% 以下，预测效果较为稳定；(2)BP-LSTM 模型和 GA-BP 模型总体预测误差在 6%~9% 之间，预测效果一般且不稳定；(3)BPNN 模型和 LSTM 模型总体误差为 8%~12%，预测效果较差。

图 7 更加直观地反映了各个模型的预测效果。GA-BP 模型和普通的 BP-LSTM 模型仅从模型角度优化，未能考虑云服务需求受时序因素的影响，难以准确把握云计算资源负荷特征，因而在预测精度和预测稳定性上存在劣势。单一的 BPNN 模型和 LSTM 模型不具备对预测结果的二次优化，且未能充分考虑对云计算资源负荷具有重要影响的时序因素，因而预测效果较差。

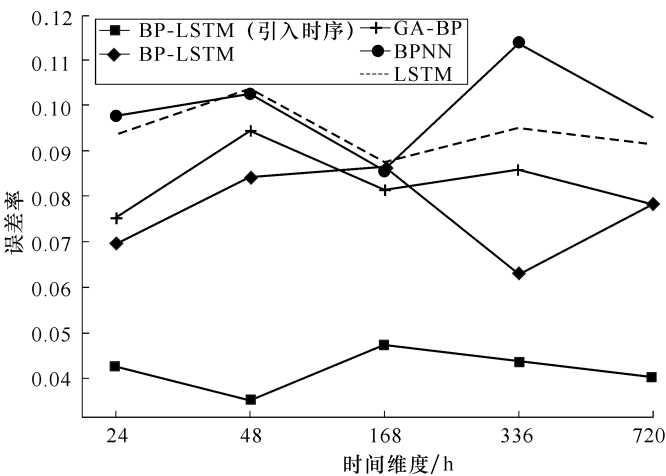


图7 各模型不同时间维度上预测的误差率对比

综合上述分析，笔者提出的引入时序因素的 BP-LSTM 组合预测模型在不同时间维度下的预测精度更高且更具稳定性。

五、结束语

针对当前云计算资源负荷预测精度低、难以准确把握历史数据时序特征的问题，提出了一种基于 BPNN-LSTM 的云计算资源负荷组合预测模型，引入时序数据构建增强数据集，采用 LSTM 网络对 BPNN 网络的预测残差进行修正。采用 Wikipedia 网站的云计算资源负荷数据进行数值实验，对不同时间维度下的负荷量进行预测。实验结果表明，笔者提出的基于 BPNN-LSTM 的云计算资源负荷组合预测方法行之有效，能够对不同时间维度下的云计算资源负荷进行精准预测。

后续研究工作主要聚焦于将所提出模型与不同类型的云计算实例相结合，在不同时间维度下对多实例的云计算资源负荷进行预测，以提高该方法的普适性。

参考文献：

[1] 徐达宇,杨善林,罗贺. 基于广义模糊软集理论的云计算资源需求组合预测研究[J]. 中国管理科学,2015,23(5): 56-64.

[2] 江伟,陈羽中,黄启成,等. 一种云环境下的主机负载预测方法[J]. 计算机科学,2018,45(6):283-287.

[3] 文静,李陶深,黄汝维. IaaS 下基于预测的弹性云服务的研究[J]. 系统工程理论与实践,2014,34(s1):263-268.

[4] RUBEN V D B, VANMECHELEN K, BROECKHOVE J. IaaS reserved contract procurement optimization with load prediction [J]. Future Generation Computer Systems, 2015, 53:13-24.

[5] CHEN J, WANG Y. A resource demand prediction method based on EEMD in cloud computing [J]. Procedia Computer Science, 2018, 131:116-123.

[6] KUMAR J, SINGH A K. Workload prediction in cloud using artificial neural network and adaptive differential evolution [J]. Future Generation Computer Systems, 2018, 81:41-52.

[7] PAULRAJ G J L, FRANCIS S A J, PETER J D, et al. A combined forecast-based virtual machine migration in cloud data centers [J]. Computers & Electrical Engineering, 2018,69(6):287-300.

[8] NHUAN T, THANG N, BINH H, et al. A multivariate fuzzy time series resource forecast model for clouds using lstm and data correlation analysis[J]. Procedia Computer Science, 2018, 126: 636-645.

[9] KUMAR J, GOOMER R, SINGH A K. Long short term memory recurrent neural network (LSTM-RNN) based workload forecasting model for cloud datacenters[J]. Procedia Computer Science, 2018, 125:676-682.

[10] 孟煜,张斌,郭军,等. 云计算环境下云服务用户并发量的区间预测模型[J]. 计算机学报,2017,40(2):378-396.

[11] 李晓峰,徐玖平,王荫清,等. BP 神经网络自适应学习算法的建立及其应用[J]. 系统工程理论与实践,2004,24

- (5):1-8.
- [12] GREFF K, SRIVASTAVA R K, KOUTNIK J, et al. LSTM: a search space odyssey [J]. IEEE Transactions on Neural Networks & Learning Systems, 2016, 28(10):2222-2232.
- [13] Jentialgo. Traces_en_wikipedia [DB/OL]. (2017-07-29) [2019-05-25]. <https://github.com/asi-uniovi/lloovia/tree/master/data/paper>.
- [14] GURLEEN K, ANJU B, INDERVEER C. An intelligent regressive ensemble approach for predicting resource usage in cloud computing [J]. Journal of Parallel and Distributed Computing, 2019, 123:1-12.
- [15] ALI A, MOSTAFA G, SAJJAD T. A learning automata-based ensemble resource usage prediction algorithm for cloud computing environment [J]. Future Generation Computer Systems, 2018, 79: 54-71.
- [16] VARGHESE B, BUYYA R. Next generation cloud computing: new trends and research directions [J]. Future Generation Computer Systems, 2018,79:849-861.
- [17] 梁天新,杨小平,王良,等. 记忆神经网络的研究与发展[J]. 软件学报,2017,28(11):2905-2924.
- [18] 严旭,李思源,张征. 基于遗传算法的 BP 神经网络在城市用水量预测中的应用[J]. 计算机科学,2016,43(11):547-550.
- [19] WANG J, PENG B, ZHANG X. Using a stacked residual LSTM model for sentiment intensity prediction [J]. Neurocomputing, 2018,332(1):93-101.

Load Forecasting for Cloud Computing Resource Based on BPNN-LSTM Composite Model

CHEN Dong-lin, LIN Wen-jian, HUANG Li-li

(Hubei Electronic Commerce and Big Data Engineering Technology Center,
Wuhan University of Technology, Wuhan 430070, China)

Abstract: Effective load forecasting is of great significance to improving the utilization efficiency of cloud computing resources, reducing the energy consumption of cloud computing centers and reducing operation, and maintenance costs of cloud service providers. Aiming to the problem of low accuracy of cloud computing resource load forecasting and inaccurate grasp of the time series characteristics of historical data, a combined forecasting model of BPNN_LSTM with time series factors is proposed. Considering the date and time factors, which have significant influence on cloud computing resource load, LSTM is used to correct the residual of BPNN. Compared with other prediction models, the proposed model has higher accuracy and stability in cloud computing resource load forecasting.

Key words: cloud computing resources; load forecasting; time series factors; BP neural network; long-term and short-term memory network