

基于CEEMDAN-ConvLSTM组合模型的云计算负载预测方法

赵鹏, 周建涛, 赵大明

引用本文

赵鹏, 周建涛, 赵大明. 基于CEEMDAN-ConvLSTM组合模型的云计算负载预测方法[J]. 计算机科学, 2023, 50(6A): 220300272-9.

ZHAO Peng, ZHOU Jiantao, ZHAO Daming. [Cloud Computing Load Prediction Method Based on Hybrid Model of CEEMDAN-ConvLSTM](#) [J]. Computer Science, 2023, 50(6A): 220300272-9.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于三支聚类的云任务优化调度](#)

Optimal Scheduling of Cloud Task Based on Three-way Clustering

计算机科学, 2022, 49(11A): 211100139-7. <https://doi.org/10.11896/jsjcx.211100139>

[云环境下可验证关键词密文检索研究综述](#)

Research on Verifiable Keyword Search over Encrypted Cloud Data:A Survey

计算机科学, 2022, 49(10): 272-278. <https://doi.org/10.11896/jsjcx.220500285>

[面向超参数估计的贝叶斯优化方法综述](#)

Survey on Bayesian Optimization Methods for Hyper-parameter Tuning

计算机科学, 2022, 49(6A): 86-92. <https://doi.org/10.11896/jsjcx.210300208>

[云环境下基于属性的多关键字可搜索加密方案](#)

Expressive Attribute-based Searchable Encryption Scheme in Cloud Computing

计算机科学, 2022, 49(3): 313-321. <https://doi.org/10.11896/jsjcx.201100214>

[区块链即服务平台关键技术及发展综述](#)

Survey on Key Techniques and Development of Blockchain as a Service Platform

计算机科学, 2021, 48(11): 4-11. <https://doi.org/10.11896/jsjcx.210500159>

基于 CEEMDAN-ConvLSTM 组合模型的云计算负载预测方法

赵 鹏¹ 周建涛^{1,2,3,4,5,6,7} 赵大明¹

1 内蒙古大学计算机学院 呼和浩特 010021

2 蒙古文智能信息处理技术国家地方联合工程研究中心 呼和浩特 010021

3 生态大数据教育部工程研究中心 呼和浩特 010021

4 内蒙古自治区云计算与服务软件工程实验室 呼和浩特 010021

5 内蒙古自治区社会计算与数据处理重点实验室 呼和浩特 010021

6 内蒙古自治区大数据分析技术工程实验室 呼和浩特 010021

7 内蒙古自治区纪检监察大数据重点实验室 呼和浩特 010021

(zp15247468996@163.com)

摘 要 随着云计算技术的快速发展,越来越多的用户选择使用云服务。负载请求与资源供应的不匹配问题日益凸显,使得用户请求无法得到及时响应,极大地影响云服务质量,实时预测负载请求,将有助于及时供应资源。针对云计算环境中的负载预测方法性能低的问题,提出了一种基于自适应噪声的完备经验模态分解和卷积长时序神经网络组合模型(CEEMDAN-ConvLSTM)的云计算负载预测方法。首先运用自适应噪声的完备经验模态(CEEMDAN)分解技术对数据序列进行分解操作,将其转换为若干个易于分析和建模的子序列;然后运用卷积长时序神经网络(ConvLSTM)预测模型对这一系列子序列进行建模预测,并采用基于多进程并行计算的研究思路,实现多序列并行预测及贝叶斯优化调参;最后将预测值综合叠加以获得整个模型的预测输出,从而实现原始复杂序列数据进行高精度预测的目标。使用 Google 集群工作负载数据集进行实验验证,实验结果表明,CEEMDAN-ConvLSTM 组合模型具有良好的预测效果,相比自回归差分移动平均模型(ARIMA)、长短期记忆网络(LSTM)和卷积长时序神经网络(ConvLSTM),所提模型预测均方根误差(RMSE)指标分别提升了 30.9%,30.1%和 22.5%。

关键词: 云计算;负载预测;卷积长时序神经网络(ConvLSTM);模态分解技术;贝叶斯优化

中图法分类号 TP311.1

Cloud Computing Load Prediction Method Based on Hybrid Model of CEEMDAN-ConvLSTM

ZHAO Peng¹, ZHOU Jiantao^{1,2,3,4,5,6,7} and ZHAO Daming¹

1 College of Computer Science, Inner Mongolia University, Hohhot 010021, China

2 National & Local Joint Engineering Research Center of Intelligent Information Processing Technology for Mongolian, Hohhot 010021, China

3 Engineering Research Center of Ecological Big Data, Ministry of Education, Hohhot 010021, China

4 Inner Mongolia Engineering Laboratory for Cloud Computing and Service Software, Hohhot 010021, China

5 Inner Mongolia Key Laboratory of Social Computing and Data Processing, Hohhot 010021, China

6 Inner Mongolia Engineering Laboratory for Big Data Analysis Technology, Hohhot 010021, China

7 Inner Mongolia Key Laboratory of Discipline Inspection and Supervision Big Data, Hohhot 010021, China

Abstract With the rapid development of cloud computing technology, more and more users choose to use cloud services, and the problem of mismatch between load requests and resource supply becomes increasingly prominent. As a result, user requests cannot be timely responded, which greatly affects the cloud service quality. Real-time prediction of load requests will help the timely supply of resources. To solve the problem of low performance of load prediction methods in the cloud computing environment, a cloud computing load prediction method based on hybrid model of complete ensemble empirical mode decomposition with adaptive noise and convolutional long short-term memory(CEEMDAN-ConvLSTM) is proposed. To begin with, the data sequence is decomposed into several sub-sequences which are easy to analyze and model. Then the convolutional long short-term memory(Conv-

基金项目:国家自然科学基金(62162046);内蒙古科技攻关项目(2021GG0155);内蒙古自然科学基金重大项目(2019ZD15);内蒙古自然科学基金(2019GG372)

This work was supported by the National Natural Science Foundation of China(62162046), Key Science-Technology Program of Inner Mongolia, China(2021GG0155), Major Program of the Inner Mongolia Natural Science Foundation of China(2019ZD15) and Natural Science Foundation of Inner Mongolia, China(2019GG372).

通信作者:周建涛(cszjtao@imu.edu.cn)

LSTM) prediction model is used to predict the series of sub-sequences. The research idea based on multi-process parallel computation is adopted to realize multi-sequence parallel prediction and Bayesian optimization parameter tuning. Finally, the prediction values are integrated and superimposed to obtain the prediction output of the whole model, to achieve the goal of high-precision prediction of the original complex sequence data. The CEEMDAN-ConvLSTM hybrid model is verified by using the Google cluster workload data set. Experiment results show that the CEEMDAN-ConvLSTM hybrid model had a good prediction effect. Compared with the autoregressive differential moving average model (ARIMA), long short-term memory network (LSTM) and the convolutional long short-term memory (ConvLSTM), the Root Mean Square Error (RMSE) increases by 30.9%, 30.1% and 22.5%, respectively.

Keywords Cloud computing, Load prediction, Convolutional long short-term memory (ConvLSTM), Modal decomposition technique, Bayesian optimization

1 引言

云计算技术快速发展,广泛应用于诸多行业。然而随着云用户的迅速增加,数据中心高能耗、高污染等问题日益突出^[1]。一方面是因为云提供商引进更大规模的服务器来满足用户的需求,造成过度供应^[2];另一方面是因为负载请求与资源供应不匹配。在大多数情况下,用户端应用程序的资源需求很少是固定不变的,它们的负载会随时间不断地变化并呈现出一定的模式^[3]。如果根据最高的负载请求提供资源,则在需求较低时会造成资源的闲置;相反,如果资源供应保持在相对较低的水平,则可能会影响服务质量^[4]。这种负载请求与资源供应不匹配的情况将造成数据中心巨大的能耗浪费。负载预测就是通过精准预测未来负载请求,及时有效地调整资源供应,在满足用户需求的同时达到自身效益的最大化。目前,在云计算领域中,负载预测在资源供应、提高服务质量、高能耗解决等场景中有着广泛的应用需求,得到了越来越多研究者的关注。

在早期研究中,研究者通过研究预测对象本身的变化规律,来推测发展趋势,如自回归滑动平均模型^[5] (ARMA)、隐马尔可夫模型^[6] (HMM)等被用于负载预测。但是由于其浅层的结构,无法获取数据集的本质特征,因此预测准确性面临瓶颈。近年来,深度学习在负载预测中逐渐成为主流方法,相比传统预测模型,深度学习模型拥有强大的泛化能力和更好的精确效果,如前馈神经网络^[7] (ANN)、长短期记忆网络^[8] (LSTM)等模型在负载预测领域都取得了很好的预测效果。但是目前深度学习模型存在训练代价大、复杂度高、精确度不稳定的问题,且不可广泛使用。

为解决上述问题,本文提出了一种基于自适应噪声的完备经验模态分解和卷积长时序神经网络组合模型 (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise and Convolutional Long Short-Term Memory, CEEMDAN-ConvLSTM) 的云计算负载预测方法。一方面,由于云计算负载具有动态性及不确定性,单一模型无法充分获取数据集的重要特征,为此,本文采取“先分解后组合”的建模框架^[9],即首先运用分解技术对复杂度较高、直接建模难度较大的序列数据进行分解操作,将其转换为若干个易于分析和建模的子序列;然后运用基于机器学习预测的方法对这一系列子序列进行建模预测;最后将预测值综合叠加以获得整个模型的预测输出,从而完成对原始复杂序列数据进行高精度预测的

目标。另一方面,ConvLSTM 是为读取二维时空数据而开发的^[10],但也适用于单变量时间序列预测,即将单变量时间序列数据构建成二维序列,从而可以通过卷积神经网络来捕捉局部数据之间的关联。ConvLSTM 模型不仅具有 LSTM 模型的时序建模能力,还能像 CNN 模型一样刻画局部特征,并且组合模型通常对复杂问题具有更强的鲁棒性。

本文的主要贡献包括:

(1) 使用 ConvLSTM 模型预测单变量时间序列数据,将单变量时间序列数据构建成二维序列,从而可以通过卷积神经网络来捕捉局部数据之间的关联。

(2) 提出基于 CEEMDAN-ConvLSTM 组合模型的负载预测方法,应用“先分解后组合”的思想,有效提高预测精度。

(3) 提出基于多进程并行计算的贝叶斯优化方法,实现多序列并行预测及优化,该方法过程简单,速度快,且优化结果好,是一种十分有效的全局优化算法。

(4) 在 Google 集群工作负载数据集上的对比实验验证了基于 CEEMDAN-ConvLSTM 组合模型的负载预测方法的有效性,可以高效准确地预测出未来某个时间段内负载的请求情况。

本文第 2 节介绍相关的研究工作;第 3 节给出模型介绍,主要包括基本方法原理和 CEEMDAN-ConvLSTM 组合模型;第 4 节介绍实验环境和实验设计;第 5 节展示实验结果;最后总结全文。

2 相关工作

目前,针对云计算环境中的资源预测问题,国内外一些学者提出了多种负载预测模型,大体上可分为两类:基于传统建模方法和基于深度学习的方法。

基于传统建模方法大多通过研究预测对象本身的变化规律,去推测发展趋势。基于传统建模方法包括卡尔曼滤波模型、隐马尔可夫模型 (HMM)、自回归差分移动平均模型 (ARIMA) 等,其中 ARIMA 模型结构简单、灵活性强,具有强大的统计属性,被广泛应用在时间序列预测研究领域。Calheiros 等^[11]使用 ARIMA 模型对维基负载数据进行了有效的预测,根据预测值实现资源动态配置,然而 ARIMA 模型的缺点也很明显,首先,它要求数据序列有着平稳的变化趋势;其次,ARIMA 只能获取线性关系,而非线性关系的预测效果较差。Panneerselvam 等^[6]提出了 HMM 模型对负载进行预测,通过给定 HMM 的模型参数和一个观测序列,计算

一系列的隐状态,使得此观测序列的出现可能最大,进而得到局部最优,但是存在训练代价大和复杂度高的问题。因此,如何在降低训练代价和使用参数数量较小的情况下提高 HMM 的预测准确性仍是一个值得改进的思考点。

基于深度学习的方法是采用一系列深度学习方法处理和分析数据序列,从而实现预测功能。深度学习是机器学习领域中一个较新的研究方向,具有多层表征的功能^[12]。基于深度学习的方法包括人工神经网络(ANN)、反向传播神经网络(BP)、循环神经网络(RNN)、时间卷积网络(TCN)、长短期记忆网络(LSTM)、极端梯度提升(XGBoost)等,其中神经网络是目前针对应用变形最多的方法。Duy 等^[7]将 ANN 应用于主机负载预测任务,其比传统分布式系统中的方法具有更好的性能。与 AR 模型相比,ANN 模型最显著的优势是非线性拟合能力,它可以捕获输入和输出值之间的潜在关系,但是由于输入固定且有限,因此无法预测长期主机负载。Pacheco-Sanchez^[13]曾提出了有关 BP 神经网络的预测方法。作为神经网络里的代表性算法,BP 神经网络对于各种输入值和输出可以建立比较有效的映射效果,但由于其本质上为梯度下降法,而它所优化的目标函数又非常复杂,因此,必然会出现“锯齿形现象”,这使得 BP 神经网络效果较差。Zhang 等^[14]利用 RNN 模型对 Google 提供的集群公开数据集进行预测,但 RNN 模型在对时间序列进行预测时存在梯度消失和梯度爆炸的问题,无法处理长距离依赖的问题。LSTM 是一种改进之后的循环神经网络,可以有效解决 RNN 无法处理长距离依赖的问题,在负载预测问题上也有广泛的应用。组合模型通常对复杂问题具有更强的鲁棒性。Lin 等^[15]提出了一种组合模型 ARIMA-LSTM 对云平台负载进行预测,验证了所提出的组合模型优于单一预测模型。Song 等^[16]提出了一种基于 CNN 和 LSTM 的资源负载预测模型,该模型深度融合了 CNN 和 LSTM 出色的特征提取能力,实现了多角度的时间序列特征提取。

综上所述,由于云数据中心负载数据序列具有动态性及不确定性,且云数据中心经常会有突发事件导致的请求峰值,因此,常规模型无法充分获取复杂数据的重要特征,预测精度较低,而现存的一些组合模型的预测精度虽然优于单一预测模型,但预测复杂数据的精度也相对较低且预测时间较长。为提高云数据中心负载预测精度,本文提出一种基于 CEEMDAN-ConvLSTM 组合模型的云数据中心负载预测方法。该方法不仅提高了云数据中心负载预测精度,同时适用于其他复杂数据序列预测。

3 模型原理

本节首先对云数据中心负载预测进行形式化定义,包括问题描述及相关符号说明;其次,介绍构成所提负载预测方法的几种基本算法;最后,详细介绍本文提出的基于 CEEMDAN-ConvLSTM 组合模型的云计算负载预测方法。

3.1 问题描述

本文主要通过给定云数据中心负载情况来预测未来云数据中心负载状况,从而及时解决负载请求与资源供应不匹配的问题。

设云数据中心资源负载数据序列为 $X(n) = \{x_1, x_2, \dots,$

$x_k, \dots, x_n\}$,其中 n 为序列长度, x_k 为 k 时刻负载情况。所研究问题可描述为,设置时间步长为 t ,从给定资源负载序列 $X(n)$ 中选取前 t 个时刻的序列数据作为预测模型输入,即 $X_t = \{x_{n-t}, \dots, x_{n-2}, x_{n-1}\}$,通过预测模型训练后便可得到下一时刻云数据中心资源负载情况。由于 CPU 利用率对云数据中心的功耗和服务质量影响较大,所以通常使用 CPU 利用率研究云数据中心资源负载情况。

3.2 基本方法原理

3.2.1 CEEMDAN 分解算法

经验模态分解(Empirical Mode Decomposition, EMD)是一种针对非平稳序列数据的自适应分解算法,它通过将原始序列数据自适应地分解为一系列具有不同特征尺度的数据序列,从而可以有效提高组合预测方法的预测精度。

但是由于云数据中心负载数据序列具有动态性及不确定性,且云数据中心经常会有突发事件导致的请求峰值,使用 EMD 分解算法往往会出现模态混叠现象,即一个本征模态函数包含了明显不同的几个模态。因此本文选取了可以很好地克服 EMD 分解算法中存在的模态混叠现象的 CEEMDAN 分解算法。

CEEMDAN 分解算法是 EMD 分解算法的一种改进算法,其原理是将原始序列数据自适应地分解为一系列震荡函数,并且在分解的过程中,根据式(1)在每个阶段都添加自适应白噪声,使重构误差趋于零^[17]。

$$IMF_{k+1}(n) = \frac{1}{I} \sum_{i=1}^I E_1 \{r_k(n) + \gamma_k E_k[\omega^i(n)]\} \quad (1)$$

其中, γ 指控制白噪声能量的参数, $\omega^i(n)$ 指满足高斯分布的噪声,IMF 指原始信号被 EMD 分解之后得到的各层信号分量, I 指样本总数, $E_k(\ast)$ 指数据序列的 k 阶 IMF 分量, $r_k(n)$ 指剩余分量。

然而,云计算负载预测要求及时性,尽管 CEEMDAN 分解算法能有效提高组合预测方法的预测精度,但由于预测序列增多反而大幅度增加了预测的时间。因此本文采用基于多进程并行计算的研究思路,将各待预测 IMF 分量分配给若干进程进行后续操作,大幅度提高计算速度,使模型更实时高效。图 1 为以预测 CPU 利用率为例,进行并行预测算法的流程图。首先,采用 CEEMDAN 分解算法将原始 CPU 利用率序列数据分解为若干个 IMF 分量,并根据 IMF 分量的数目创建相同数量的进程;其次,为各进程配置预测功能所需的资源,并将各待预测 IMF 分量分配给各进程进行预测及优化;最后,待所有进程执行结束后,收集各进程预测结果。所以,多进程并行计算可以有效消除基于 CEEMDAN 分解算法的组合预测方法耗时的弊端。并行预测算法如算法 1 所示。

算法 1 并行预测算法

Input:当前一段时间内云数据中心资源负载数据序列 $X(n)$

Output:未来一段时间内云数据中心资源负载请求状况 $pre_X(m)$

1. 依据式(1),将 $X(n)$ 自适应地分解为一系列具有不同特征尺度的 IMF 分量,并计算分量个数(包括最后的剩余分量)为 $count_decom$ 。
2. 创建 $count_decom$ 个进程,并将步骤 1 中分解得到的分量逐一分配给各进程,使之——对应。
 - 2.1. 并行执行各进程。

- 2.2. 对于每个进程,分别构建 ConvLSTM 网络,并给定网络参数范围,以均方根误差(RMSE)为评判标准,利用贝叶斯优化函数(BayesianOptimization)求出 RMSE 值最小的参数组合。
- 2.3. 使用步骤 2.2 求出的最优参数组合配置 ConvLSTM 网络,并对 IMF 分量进行预测。
- 2.4. 保存各进程的预测数据。
3. 待以上所有进程执行结束,收集各进程的预测数据,依次相加得到最终的预测结果 pre_X(m)。

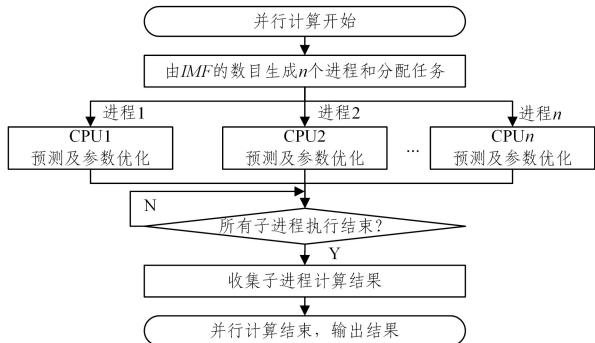


图 1 并行预测流程图
Fig. 1 Flowchart of parallel prediction

3.2.2 ConvLSTM 模型

ConvLSTM 模型是 LSTM 模型的一种改进模型,最早由 Shi 等^[18]在短时降水预测中提出,用于处理时空预测问题。ConvLSTM 模型不仅具有 LSTM 模型的时序建模能力,而且还能像 CNN 模型一样刻画局部特征。因此本文使用 ConvLSTM 模型预测云数据中心资源负载请求。

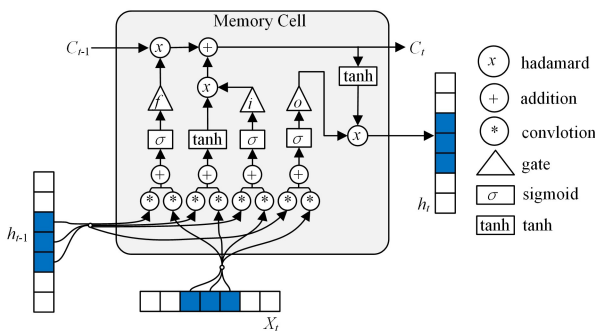


图 2 ConvLSTM 网络结构
Fig. 2 Architecture of ConvLSTM network

但是,ConvLSTM 是为读取二维时空数据而开发的,而云数据中心负载数据序列为单变量时间序列数据,为此本文将单变量时间序列数据构建二维序列,从而可以通过 ConvLSTM 模型来捕捉局部数据之间的关联。ConvLSTM 网络结构如图 2 所示。ConvLSTM 模型是将 LSTM 模型的前馈方法从 Hadamard 乘积变为卷积。

ConvLSTM 模型虽具有强大的非线性泛化能力和自动特征提取能力,但要真正运用到云计算负载预测中,模型优化更为关键,不仅要求保证模型预测精度,还要求保证时间效率。然而,常规优化方法搜索速度较慢,且容易陷入局部最优,因此本文选用贝叶斯优化算法。

贝叶斯优化是一种十分有效的全局优化算法^[19],只需经过少次数目标函数评估就可获得复杂目标函数的最优解,适用于深度学习模型的超参数调优问题。其原理是给定优化的目标函数,无需知道内部结构以及数学性质,只需指定输入和输出即可,通过不断地添加样本点来更新目标函数的后验

分布,直到后验分布基本贴合于真实分布,便可得到模型的最优超参数组合。基于贝叶斯优化的 ConvLSTM 预测算法如算法 2 所示。因为 ConvLSTM 网络的初始权值和阈值是随机产生的,即使找到最优参数组合,网络每次结果仍不同,所以本文使用阈值约束方法优化网络预测能力,通过循环执行目标 ConvLSTM 网络,直到网络预测能力达到理想预测效果阈值。然而,云计算负载预测要求及时性,因此不能无限循环 ConvLSTM 网络,所以本文通过限制迭代次数去约束优化时间,一方面提高了 ConvLSTM 网络预测效果,另一方面保证了 ConvLSTM 网络的时效性。

算法 2 基于贝叶斯优化的 ConvLSTM 预测算法

Input: 当前一段时间内云数据中心资源负载数据序列 X(n)

Output: 未来一段时间内云数据中心资源负载请求状况 pre_X(m)

1. 数据集预处理,并划分训练集和验证集。
2. 构建 ConvLSTM 网络。
3. 给定 ConvLSTM 网络参数范围,以均方根误差(RMSE)为评判标准,利用贝叶斯优化函数(BayesianOptimization)求出 RMSE 值最小的参数组合。
4. 依据经验设定理想预测模型 RMSE 上限值 Max_RMSE 及可接受的迭代次数 num。
5. For i in num:
 - 5.1. 使用步骤 3 求出的最优参数组合配置 ConvLSTM 网络,并验证网络预测效果,求出当前 RMSE 值。
 - 5.2. if RMSE 大于 Max_RMSE 且当前迭代次 i 小于 num,则重复步骤 5.1; 否则停止循环,输出此时 ConvLSTM 网络预测结果。

3.3 CEEMDAN-ConvLSTM 组合模型

云数据中心负载数据序列具有动态性及不确定性,且云数据中心经常会有突发事件导致的请求峰值,因此,单一模型无法充分获取复杂数据的重要特征,预测精度较低。本文根据 CEEMDAN 分解算法对数据的分解特性和 ConvLSTM 模型强大的特征提取功能,提出了一种基于 CEEMDAN-ConvLSTM 组合模型的云数据中心负载预测方法。模型结构如图 3 所示。首先,采用 CEEMDAN 算法对输入的负载序列数据进行处理,分解得到若干个子序列,可以更全面地掌握负载序列数据特征,便于建立更精准的负载预测模型;其次,采用经贝叶斯优化的 ConvLSTM 神经网络进行子序列预测,实现自动优化,降低网络参数选择的随机性,提高 ConvLSTM 神经网络的预测能力。本小节将对提出的基于 CEEMDAN-ConvLSTM 组合模型的云计算负载预测方法的实现过程进行详细介绍。

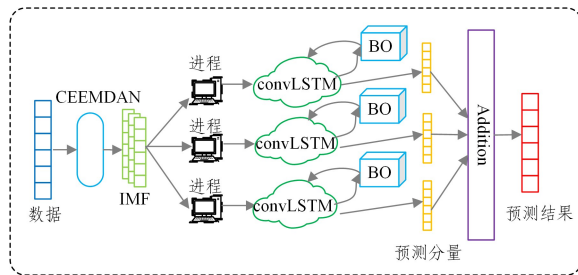


图 3 模型结构
Fig. 3 Model structure

第一阶段 采用 CEEMDAN 算法对输入的负载序列数据进行处理,将其转换为若干个易于分析和建模的模式分量(IMF)和 1 个残差分量。

第二阶段 采用基于多进程并行计算的研究思路,将各待预测 IMF 分量分配给若干进程进行后续操作,大幅度提高计算速度,使模型更实时高效,并行预测算法如算法 1 所示。

第三阶段 将各阶次 IMF 分量和残差分量输入 ConvLSTM 神经网络中,将单变量时间序列数据构建成二维序列,从而捕捉局部数据之间的关联,实现数据预测;并采用贝叶斯优化的方法进行自动调参,基于贝叶斯优化的 ConvLSTM 预测算法如算法 2 所示。

第四阶段 将各阶次 IMF 分量和残差分量的预测结果依次相加得到最终的预测结果。

4 实验设置

4.1 实验平台

为了实现负载预测模型及评估其性能,本文在 PyCharm 2020.3.3 x64 中利用 Python 编程语言实现,并使用 TensorFlow-2.4.1 框架实现所提模型和对比模型,实验过程在个人台式电脑上完成,其配置如表 1 所列。

表 1 硬件配置及系统参数

Table 1 Hardware configuration and system parameters

CPU	RAM	Type of OS	OS
Intel(R)Core(TM) i7-8700	16 GB	64 Bit	Windows 10
CPU@3.20GHz			

4.2 数据集

本文使用 Google 集群工作负载数据集^[20],该集群包含超过 2500 万个任务,分布在 12500 多台主机上,记录了 29 天内约 12500 个计算节点的使用状态,收集的信息包括 CPU 利用率、内存利用率、磁盘 I/O 时间、磁盘空间等。本文实验选取该数据集中一台主机在 3 天内的使用状态进行研究,节点间隔时间为 5 min,共 835 条数据。使用 CPU 利用率进行预测分析,并将选取的数据集划分为两组,分别为用于计算模型权重及选择超参数的训练集和用于评估模型效果的测试集,并选取前 75% 作为训练集,后 25% 作为测试集。

4.3 评价指标

为了检验各模型预测效果的有效性,本文选择了 3 个不同的评价指标,分别为均方根误差 (RMSE)、平均绝对误差 (MAE)、拟合程度 (R-squared)。其中均方根误差用来衡量预测值和真实值之间的误差;平均绝对误差用来反应预测值误差的实际情况;拟合程度用来反应模型的拟合度,其值越接近 1,说明模型拟合度越好。计算式如下:

$$RMSE = \sqrt{\frac{1}{S} \sum_{i=1}^S (y_i - \hat{y}_{i-1})^2} \quad (2)$$

$$MAE = \frac{1}{S} \sum_{i=1}^S |y_i - \hat{y}_{i-1}| \quad (3)$$

$$R\text{-squared} = 1 - \frac{\sum_{i=1}^S (y_i - \hat{y}_{i-1})^2}{\sum_{i=1}^S (y_i - \bar{y})^2} \quad (4)$$

其中, S 为样本的个数, y_i 指样本实际值, \hat{y}_{i-1} 指样本预测值, \bar{y} 指样本的均值。

由式(2)一式(4)可得,预测精度越高, $|y_i - \hat{y}_{i-1}|$ 的值越小, RMSE 和 MAE 的值越小, R-squared 的值越大。

4.4 对比方法

本文对比了以下两类预测方法。

第一类是基于单一模型的预测方法。

(1) ARIMA^[11]: 通过过去观测变量和当前观测变量的值,结合随机扰动因素进行回归建立模型描述。

(2) LSTM^[8]: 通过引入记忆单元避免 RNN 模型所存在的梯度消失或者梯度膨胀的问题,解决长期依赖问题。

(3) XGBoost^[21]: 通过学习新函数去拟合前一次预测的残差,并根据数据序列的特征去计算出每个节点对应的分数,累加所有的分数就是该数据序列的预测值。

(4) ConvLSTM^[10]: 不仅具有 LSTM 的时序建模能力,而且还能像 CNN 一样刻画局部特征,可以说是时空特性兼备。

第二类是基于组合模型的预测方法。

(1) CNN-LSTM^[16]: 使用 CNN 提取时间序列的短时序依赖特性,使用 LSTM 提取时间序列的长时间依赖特征,从而实现高质量的长期时序预测。

(2) EMD-LSTM^[22]: 在使用 LSTM 模型预测前,先对输入数据进行 EMD 分解,从而实现对复杂多变数据进行高质量预测。

(3) EMD-ConvLSTM^[23]: 具有强大的非线性泛化能力以及同时提取时序空间依赖信息。

5 实验结果

模型预测结果如图 4 和图 5 所示,图 4 为运用 CEEMDAN 分解算法对原始数据序列进行处理得到的各阶次 IMF 分量曲线图,其中红色曲线为原始 CPU 数据序列,绿色曲线为各阶次 IMF 分量,且各 IMF 分量叠加起来就是原始序列。

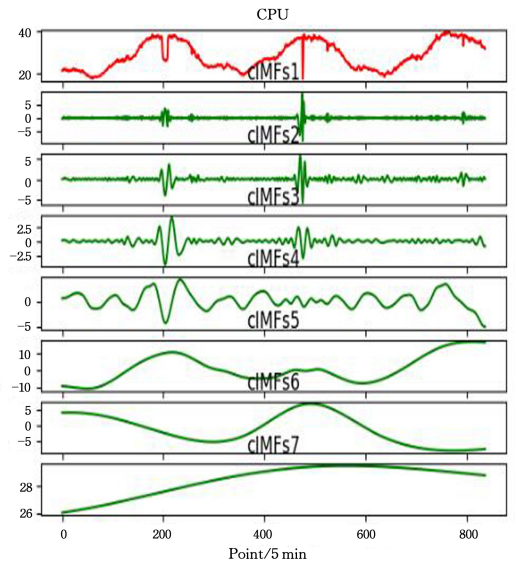


图 4 各阶次 IMF 分量的曲线图 (电子版为彩图)

Fig. 4 Curve of each IMF components

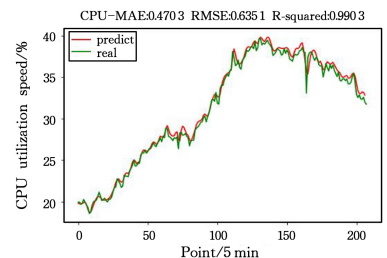


图 5 预测曲线图

Fig. 5 Curve of prediction

图 5 为本文提出的基于 CEEMDAN-ConvLSTM 组合模型的负载预测方法的预测效果图,绿色曲线代表原始数据序列,红色曲线代表预测数据序列,图形上方为评价指标值,且模型预测精度越高, RMSE 和 MAE 的值越小, R-squared 的值越大,可直观地发现本文模型具有良好的预测效果。

5.1 不同预测方法对比分析

为了验证本文提出的基于 CEEMDAN-ConvLSTM 组合模型的负载预测方法的预测效果,将其与前文列举的对比模型进行了对比实验,表 2 列出了不同预测模型在同一数据集上的预测结果对比,同时为了避免实验的偶然性,本文重复实验 20 次,对评价指标取均值。图 6 和图 7 分别为单一预测模型和组合预测模型与本文提出的基于 CEEMDAN-ConvLSTM 组合模型的负载预测方法的预测效果的对比图,为了凸显各模型预测效果差异,本文将预测效果图分为 4 段展示。从表 2 可以看出,本文提出的负载预测方法在 RMSE, MAE 和 R-squared 指标上优于所有对比模型,相比于 ARIMA, LSTM, XGBoost 和 ConvLSTM, CEEMDAN-ConvLSTM 模型

的 RMSE 指标分别提升了 30.9%, 30.1%, 32.9% 和 22.5%, 还验证了组合模型的预测精度在一定程度上是优于单一预测模型的。

表 2 各模型预测评价

Table 2 Valuation for each prediction model

预测模型	评价指标		
	RMSE	MAE	R-squared
ARIMA	0.9197	0.6651	0.9797
LSTM	0.9088	0.6804	0.9802
XGBoost	0.9474	0.7196	0.9780
ConvLSTM	0.8197	0.6045	0.9835
CNN-LSTM	0.8398	0.6226	0.9818
EMD-LSTM	0.8747	0.6414	0.9812
EMD-ConvLSTM	0.7456	0.5126	0.9867
CEEMDAN-ConvLSTM	0.6351	0.4703	0.9903

通过图 6 和图 7,可以发现各预测模型的预测能力都非常不错,而本文提出的负载预测方法始终略优于其他对比模型的预测能力。

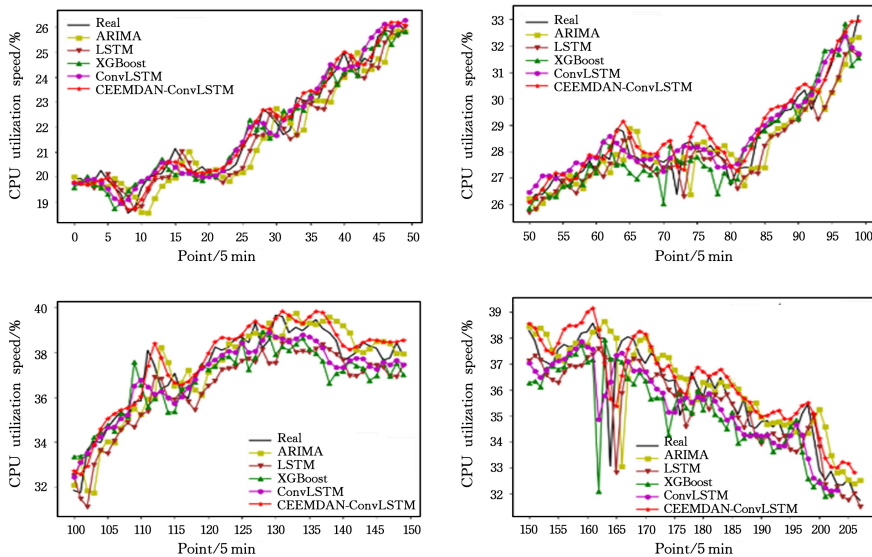


图 6 各模型预测曲线图

Fig. 6 Prediction curve of each model

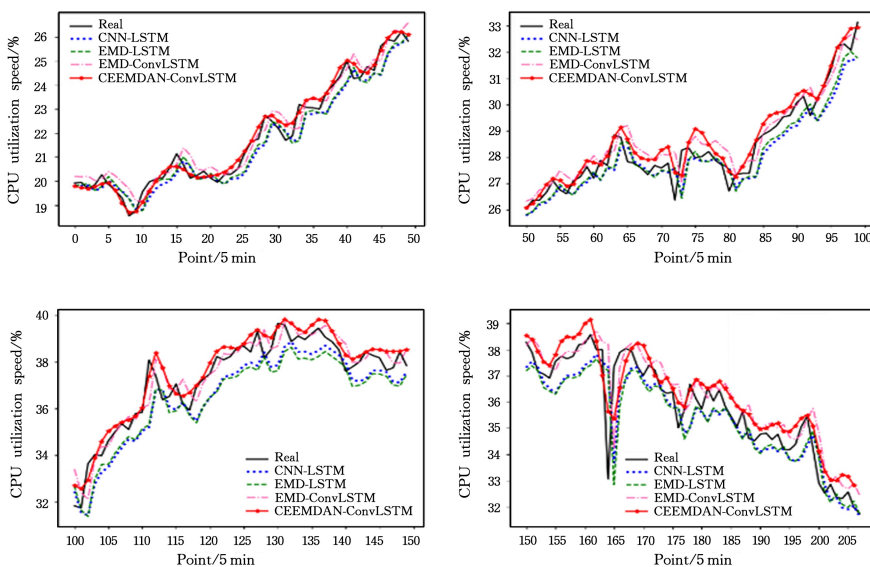


图 7 各模型预测曲线图

Fig. 7 Prediction curve of each model

一方面是因为本文提出的负载预测方法的预测曲线更接近于真实曲线,拟合度达到 99.03%,而且仔细观察图 6 和图 7,可以发现本文提出的负载预测方法可以很精确地捕捉局部数据之间的关联关系,不受特殊数据影响,比如极增极减会影响后续预测方向;另一方面是因为本文提出的负载预测方法对于云数据中心负载预测更具有实用价值,由于负载预测具有特殊性,一般要求负载预测模型不仅具有高预测精度,还要有很好的应对极端情况的能力,比如极增需要提前告警,极减需要降耗,而本文提出的负载预测方法恰恰符合。从图 6 和图 7 可以发现,对比模型预测值一般低于真实值,尤其是在极值的情况下,不能有效告警或者大幅度调整服务器去降耗,影响了服务质量,而本文所提出的负载预测方法的预测值一般略高于真实值,可以有效应对各种情况和减缓供应压力。

为了进一步验证本文提出的基于 CEEMDAN-ConvLSTM 组合模型的负载预测方法优势的有效性,本文进行了差异显著性分析。以 CNN-LSTM 模型为例,表 3 列出了本文所提出的负载预测方法和 CNN-LSTM 模型在同一数据集下以 RMSE 为评价指标的 6 对测试结果,对表 3 数据进行差异显著性分析,分析结果如表 4 所列,通常,从 F 、 P -value、 F_{crit} 这三个值便可得出差异显著性,如果 F 大于 F_{crit} ,则已经比出对差异,再结合 P -value,若 P -value 大于 0.01、小于 0.05,表示差异显著;若 P -value 小于 0.01,则表示差异极显著。如果 F 小于 F_{crit} ,那么 P -value 肯定高于 0.05,则表示两组数据无差异。从表 4 可以发现 F 远大于 F_{crit} ,且 P -value 等于 0.00017,很显然两种方法的测试结果存在显著性差异,同理,对其他对比模型与本文所提出的负载预测方法进行差异显著性分析,结果仍相同,最终可以得出结论:本文所提出的基于 CEEMDAN-ConvLSTM 组合模型的负载预测方法的预测效果均比其他预测模型取得显著提升。

表 3 预测结果

Table 3 Prediction results

编号	RMSE	
	CEEMDAN-ConvLSTM	CNN-LSTM
1	0.5184	0.7615
2	0.6351	0.7983
3	0.6953	0.8398
4	0.7095	0.9500
5	0.7122	0.9863
6	0.5637	0.8271

表 4 差异显著性分析结果

Table 4 Significant difference analysis results

差异源	SS	df	F	P-value	F _{crit}
列	0.14714	1	100.2482	0.00017	6.60789

5.2 组成模型对比分析

ConvLSTM 是将 CNN 卷积操作作为 LSTM 的一个重要组成部分,并应用于每个时间步,用卷积运算代替 LSTM 单元中每个门的矩阵乘法,可以有效捕捉局部数据之间的关联。相比于 CNN-LSTM,CNN-LSTM 仅对输入数据序列进行卷积操作,而 ConvLSTM 不仅对输入数据序列进行卷积操作,还增加了对前一时刻输出进行卷积操作,提升了记忆单元的记忆和遗忘能力,从而提高了模型学习长期依赖关系的能力、

自动特征提取能力和非线性泛化能力。表 5 列出了 LSTM,ConvLSTM,CNN-LSTM 3 种模型在同一数据集上的预测结果对比,很明显 ConvLSTM 在 RMSE,MAE 和 R-squared 指标上优于其他两种模型,相比 LSTM 和 CNN-LSTM,ConvLSTM 模型的 RMSE 指标分别提升了 9.8% 和 2.4%,验证了 ConvLSTM 模型具有更强的学习长期依赖关系的能力、自动特征提取能力和非线性泛化能力。

表 5 基于 LSTM 的组成模型预测评价

Table 5 Composition model prediction evaluation based on LSTM

预测模型	评价指标		
	RMSE	MAE	R-squared
LSTM	0.9088	0.6804	0.9802
ConvLSTM	0.8197	0.6045	0.9835
CNN-LSTM	0.8398	0.6226	0.9818

为了验证本文采取“先分解后组合”的建模框架的有效性,本文对 ConvLSTM,EMD-ConvLSTM,CEEMDAN-ConvLSTM 这 3 种预测模型进行了对比实验,实验结果如表 6 所列。从表 6 可以发现 EMD-ConvLSTM 和 CEEMDAN-ConvLSTM 两种预测模型的预测精度均高于 ConvLSTM,说明应用“先分解后组合”的思想,可以提高预测精度。而且 CEEMDAN-ConvLSTM 的预测精度略优于 EMD-ConvLSTM,原因是 CEEMDAN 可以很好地克服 EMD 算法中存在的模态混叠的问题,验证了本文使用 CEEMDAN 分解算法的合理性。

表 6 基于分解算法的组成模型预测评价

Table 6 Composition model prediction evaluation based on decomposition algorithm

预测模型	评价指标		
	RMSE	MAE	R-squared
ConvLSTM	0.8197	0.6045	0.9835
EMD-ConvLSTM	0.7456	0.5126	0.9867
CEEMDAN-ConvLSTM	0.6351	0.4703	0.9903

5.3 优化方法对比分析

模型优化是模型构建的必经之路,也是其中的重要组成部分。本文采用贝叶斯优化的方法进行调参,采用基于多进程并行计算的研究思路来实现模型,因为贝叶斯优化的方法过程简单,速度快,且优化结果好,是一种十分有效的全局优化算法,并且采用基于多进程并行计算的建模方法可以有效减少模型运行时间,增强模型的实用性。表 7 列出了贝叶斯优化方法与其他两种典型优化算法特点的比较。网格搜索和遗传算法一般根据邻域信息或者通过选择、交叉、变异等操作模拟生物进化和群体智能过程去探索最优解,实现过程较复杂,搜索速度较慢,且容易陷入局部最优。相比上述优化算法,贝叶斯优化方法不仅可以提升求解效率,还减少了评估代价,只需经过较少次目标函数评估就可获得复杂目标函数的最优解^[24]。一方面是因为贝叶斯优化方法引入了主动选择策略,从而减少了不必要的目标函数评估,有效提升了求解效率;另一方面贝叶斯优化为参数引入不确定性,考虑参数的先验分布,通过平均得到参数,因此相比使用最大似然估计拟合参数等方法,该方法不易发生过拟合^[18]。因此本文选用贝叶斯优化的方法进行调参。

表7 几种优化算法特点对比

Table 7 Comparison of characteristics of several optimization algorithms

优化算法	利用 先验知识	参数引入 不确定性	主动 选择策略	最小 代价
网格搜索	×	×	×	×
遗传算法	×	×	×	×
贝叶斯优化	√	√	√	√

表8列出了本文所提出的负载预测方法顺序执行与并行执行的运行时间比较,第一行代表不涉及参数优化的预测模型的预测执行时间,第二行代表涉及参数优化的预测模型的预测执行时间。从表8可以发现采用顺序执行的时间是采用并行执行时间的5倍。因此,采用基于多进程并行计算的建模方法可以有效减少模型运行时间,而采用基于多进程并行计算的建模方法会增加模型的复杂度。虽然不涉及参数优化时的顺序执行时间在可接受范围内,但是,深度学习模型参数并不固定,受输入数据影响,每次执行都需要参数优化,而涉及参数优化时顺序执行时间远远超过可接受范围,预测结果已无实用价值,因此采用基于多进程并行计算的建模方法还是有必要的。

表8 模型执行时间对比

Table 8 Comparison of model execution time

模型	用时/s	
	顺序	并行
CEEMDAN-ConvLSTM	108.5993	14.9337
BO-CEEMDAN-ConvLSTM	1008.9683	295.5669

结束语 为解决负载请求与资源供应不匹配导致的能耗问题,本文提出了一种基于自适应噪声的完备经验模态分解和卷积长时序神经网络组合模型(CEEMDAN-ConvLSTM)的云计算负载预测方法,并采用基于多进程并行计算的贝叶斯优化方法,实现多序列并行预测及优化。使用Google集群工作负载数据集进行实验验证,实验结果表明,本文提出的负载预测方法具有良好的预测效果,同时,在RMSE, MAE和R-squared指标上优于其他预测模型,表明本文提出的负载预测方法可以高效准确地预测未来某个时间段内负载的请求情况,及时地解决应用需求大幅度增减的状况,从而有效地为用户提供可靠的云计算服务。

参考文献

- [1] AN C Y. Research on Cloud Application Load Modeling and Resource Prediction Method[D]. Hohhot: Inner Mongolia University, 2018.
- [2] WANG B, LIU F, LIN W. Energy-efficient VM scheduling based on deep reinforcement learning[J]. Future Generation Computer Systems, 2021, 125: 616-628.
- [3] ZHAO D M, ZHOU J T, LI K. An energy-aware algorithm for virtual machine placement in cloud computing[J]. IEEE Access, 2019, 7: 55659-55668.
- [4] SU M F, WANG G J, LI R F. Resource Deployment and Task Scheduling Optimization based on Prediction in EdgeCloud Collaborative Computing[J]. Journal of Computer Research and Development, 2021, 58(11): 2558-2570.
- [5] PANNEERSELVAM J, LIU L, ANTONOPOULOS N. InOt-RePCoN: Forecasting user behavioural trend in large-scale cloud environments[J]. Future Generation Computer Systems, 2018, 80: 322-341.
- [6] PANNEERSELVAM J, LIU L, ANTONOPOULOS N, et al. Workload analysis for the scope of user demand prediction model evaluations in cloud environments[C] // 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing. IEEE, 2014: 883-889.
- [7] DUY T V T, SATO Y, INOBUCHI Y. Improving accuracy of host load predictions on computational grids by artificial neural networks[J]. International Journal of Parallel, Emergent and Distributed Systems, 2011, 26(4): 275-290.
- [8] LIU Y. Container Load Prediction and Integration Strategy in Cloud Environment[D]. Guangzhou: South China University of Technology, 2020.
- [9] CHENG X L. Research on Time Series Prediction Based on Empirical Mode Decomposition[D]. Dalian: Dalian University of Technology, 2018.
- [10] SUN Z, ZHAO M. Short-term wind power forecasting based on VMD decomposition, ConvLSTM networks and error analysis[J]. IEEE Access, 2020, 8: 134422-134434.
- [11] CALHEIROS R N, MASOUMI E, RANJAN R, et al. Workload prediction using ARIMA model and its impact on cloud applications' QoS[J]. IEEE Transactions on Cloud Computing, 2014, 3(4): 449-458.
- [12] LIU J W, WANG Y F, LUO X L. Advances in deep memory Networks[J]. Chinese Journal of Computers, 2021, 44(8): 41.
- [13] PACHECO-SANCHEZ S, CASALE G, SCOTNEY B, et al. Markovian workload characterization for QoS prediction in the cloud[C] // 2011 IEEE 4th International Conference on Cloud Computing. IEEE, 2011: 147-154.
- [14] ZHANG W, LI B, ZHAO D, et al. Workload prediction for cloud cluster using a recurrent neural network[C] // 2016 International Conference on Identification, Information and Knowledge in the Internet of Things(IKI). IEEE, 2016: 104-109.
- [15] LIN T, FENG J K, HAO Z X, et al. Research on cloud computing resource load prediction based on combination prediction model[J]. Computer Engineering and Science, 2020, 42(7): 1168-1173.
- [16] SONG J, ZHANG L, XUE G, et al. Predicting hourly heating load in a district heating system based on a hybrid CNN-LSTM model[J]. Energy and Buildings, 2021, 243: 110998.
- [17] ZHANG P, WANG M. Variation Characteristics Analysis and Short-Term Forecasting of Load Based on CEEMDAN[C] // 2021 International Symposium on Electrical, Electronics and Information Engineering. 2021: 493-500.
- [18] SHI X J, CHEN Z, WANG H, et al. Convolutional LSTM network: A machine learning approach for precipitation now-casting[C] // Advances in Neural Information Processing Systems. 2015: 802-810.
- [19] CUI J X, YANG B. Bayesian optimization algorithm and applica-

- tion review[J]. Journal of software, 2018, 29(10): 3068-3090.
- [20] REISS C, WILKES J, HELLERSTEIN J. Google cluster usage traces: format + schema[EB/OL]. <https://github.com/google/cluster-data>.
- [21] LIAO X, CAO N, LI M, et al. Research on short-term load forecasting using XGBoost based on similar days[C]// 2019 International Conference on Intelligent Transportation, Big Data & Smart City(ICITBS). IEEE, 2019: 675-678.
- [22] YU J, DING F, GUO C, et al. System load trend prediction method based on IF-EMD-LSTM[J]. International Journal of Distributed Sensor Networks, 2019, 15(8): 15501477198.
- [23] FU W, WANG K, TAN J, et al. A composite framework coupling multiple feature selection, compound prediction models and novel hybrid swarm optimizer-based synchronization optimization strategy for multi-step ahead short-term wind speed forecasting[J]. Energy Conversion and Management, 2020, 205: 112461.
- [24] JIANG Y Z, HAO Z F, ZHANG Y S, et al. Bayesian predictive evolutionary algorithm[J]. Chinese Journal of Computers, 2014, 37(8): 1846-1858.



ZHAO Peng, born in 1998, postgraduate. His main research interests include cloud computing and machine learning.



ZHOU Jiantao, born in 1972, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. Her main research interests include formal methods, cloud computing and software engineering.