

doi:10.11835/j.issn.1000-582X.2020.303

# 基于改进深度森林的短期电力负荷预测模型

彭飞<sup>1</sup>, 马煜<sup>2</sup>, 张晓华<sup>1</sup>, 吴奕<sup>3</sup>, 邓文琛<sup>3</sup>, 陈志奎<sup>4</sup>

(1. 国家电网公司东北分部, 沈阳 110180; 2. 国网辽宁沈阳供电公司, 沈阳 110811; 3. 中国科学院沈阳计算技术研究所有限公司, 沈阳 110168; 4. 大连理工大学软件学院, 辽宁大连 116024)

**摘要:**深度学习模型通过学习数据的深层特征能够有效提高电力负荷预测的准确率,但同时也带来了超参数较多、模型可解释性差等问题。针对这些问题,文中将深度森林模型引入短期电力负荷预测领域。在多粒度级联森林模型的基础上改进了多粒度窗口扫描方法,调整窗口大小与滑动步长,使模型能够在不同时间尺度下提取电力负荷数据的周期性特征。此外,改进深度森林输出层的计算方法,将输出结果由离散的类向量改进为连续的预测值,进而提高模型的精确度。最后在中国东北电网的实测数据中验证了文中所提出方法的可行性与有效性。从实验结果可知,改进深度森林算法在较高预测精度的情况下能取得更高的准确率,并且相较于深度神经网络具有更快的学习速度。

**关键词:**电力负荷预测;深度森林;预测方法;智能电网

**中图分类号:**TP391

**文献标志码:**A

**文章编号:**1000-582X(2022)05-001-008

## Short term power load forecasting model based on improved deep forest

PENG Fei<sup>1</sup>, MA Yu<sup>2</sup>, ZHANG Xiaohua<sup>1</sup>, WU Yi<sup>3</sup>, DENG Wenchen<sup>3</sup>, CHENG Zhikui<sup>4</sup>

(1. Northeast Branch of State Grid Corporation of China, Shenyang 110180, P. R. China; 2. State Grid Shenyang Electric Power Supply Company, Shenyang 110811, P. R. China; 3. Shenyang Institute of Computing Technology Co. Ltd., Chinese Academy of Sciences, Shenyang 110168, P. R. China; 4. School of Software Technology, Dalian University of Technology, Dalian, Liaoning 116024, P. R. China)

**Abstract:** Deep learning method can help to learn the deep features of power load data and improve the accuracy of prediction, but it also brings problems, such as large amounts of super parameters and poor interpretability of the model. To solve these problems, this paper introduces the deep forest model for short-term load forecasting. Based on the multi-Grained Cascade forest model, the multi-granularity window scanning method is improved with adjusted window size and sliding step size, so that the model can extract the periodicity characteristics of power load data in different time scales. In addition, the calculation method of deep forest output layer is improved with changing the output result from discrete class vector to continuous predicted value, improving the accuracy of the model. Finally, the feasibility and effectiveness of the proposed method are verified with the measured data of northeast China power grid. The

收稿日期:2020-02-15 网络出版时间:2020-04-22

基金项目:国家自然科学基金资助项目(61672123)。

Supported by National Natural Science Foundation of China(61672123).

作者简介:彭飞(1972—),男,硕士,主要从事研究电力系统及其自动化研究,(E-mail)feipeng209@163.com。

通讯作者:陈志奎,男,博士,教授,主要从事为大数据计算、物联网和人工智能研究,(E-mail)zkchen@dlut.edu.cn。

experimental results show that the improved deep forest algorithm can achieve higher accuracy with higher prediction accuracy, and has faster learning speed than the deep neural network.

**Keywords:** power load forecasting; deep forest; forecasting method; smart grid

电力负荷预测主要是指结合历史数据和经济社会发展因素等对未来某一个时段的电力运行负荷、用电市场需求、用电发展形势、用电量等因素进行综合的预测及数值推算。负荷预测的研究对象是不确定事件,虽然整体上随着居民和企业的生产活动显现周期性变化,但同时也受到气温、降水、湿度、节假日、市场交易等各种环境因素的制约显现随机性变化。这种不确定性对负荷预测研究和技术提出了精细化、综合化和多元化发展的需求。因此,添加环境因素、引入新的算法、优化模型结构进而提高负荷预测精度是电力系统稳定、安全、经济运行的关键。

近年来,随着算法的优化更新和机器学习方法的应用,电力负荷预测技术也进一步提升。但传统的统计学方法和机器学习中的经典算法多属于浅层学习方法,能够拟合的数据量较小,学习能力较差,导致预测结果的误差较大,难以满足越来越高的预测精度需求。随着深度学习模型的不断发展,采用深度模型进行电力负荷预测的研究也越来越多,在短期电力负荷预测方面取得了许多突破。但深度模型往往需要使用大量的超参数,算法参数设定多为长期尝试结果或长期累积经验,人为干预较大,收敛速度较慢。因此,文中将深度森林算法引入电力负荷预测领域,解决深度模型参数多、收敛速度较慢等问题;针对电力预测数据的特点对模型进行改进,提出改进后的深度森林算法(DF, deep forest),以此提升预测效果;并采用中国东三省及内蒙古东部电网的实际运行的数据进行验证,证明该方法的可行性。

## 1 研究现状

电力负荷预测的目标可分为三类:超短期预测、短期预测与中长期预测,文中主要研究其中的短期预测方法。传统的短期预测方法主要是回归分析法,该方法利用大量历史数据分析电力负荷的波动规律,建立拟合函数,从而进行历史、现在、未来电力负荷发展方式的一致性预测<sup>[1-2]</sup>。孙建梅等<sup>[3]</sup>在大量使用灰色模型的基础上增加数据预处理工作,并对电力负荷进行了中长期预测,扩展了模型的应用范围。张志等<sup>[4]</sup>通过关联矩阵筛选得到影响负荷变化的强相关因素并对其按季节进行分解,在此基础上利用时滞效应检验和主成分分析法提升数据质量,取得了较高的预测精度。但回归分析法难以确定回归线类型、回归变量,难以达到理论上的最佳预测值<sup>[5]</sup>。

随着机器学习的发展,大量机器学习方法也被应用于短期负荷预测领域,包括支持向量机(SVM, support vector machine)、灰色模型、指数平滑法等。研究发现,神经网络法对短期的电力负荷预测更准确<sup>[6]</sup>。简献忠等<sup>[7]</sup>构建了一种全过程优化的支持向量机模型进行短期负荷预测,较之前的局部优化模型能够取得更高的预测准确度。

但这些机器学习方法大多还停留在浅层学习,对于数据特征学习能力较差,能够拟合的数据量较小。因此,随着算法技术的发展,大量深度学习模型也被应用于电力负荷短期预测。张静等<sup>[8]</sup>采用了基于拥挤距离法改进的多目标粒子群优化算法和 BP 神经网络算法对电力负荷预测研究,得到的预测精度更高、相对误差较小。文献[9-10]利用长短记忆网络(long-short term memory)来解决深度学习模型中常见的梯度消失的问题,并在此基础上对个人住宅的负荷进行预测,取得了较好的效果。Chen 等<sup>[11]</sup>将数值预测问题转化为图像处理任务,使用卷积神经网络(convolutional neural networks)模型输入数据执行精确的聚类,考虑了温度、湿度、风速等外部影响因素,具有很高的精度。

但是深度模型也受限于较大的数据需求,在数据量较小时难以使用。此外,深度学习模型的“黑箱”性质使得其可解释性较差,超参数设置较多,在许多领域并不被广泛认可。针对这些问题,文章将深度森林模型引入短期电力负荷预测领域。深度森林模型<sup>[12]</sup>是一种 2017 年提出来的深度学习模型,它的核心思想是对随机森林进行集成。相较于其他深度学习模型,深度森林方法在每一层自主训练参数,不需要反向传播过程,因此对于训练集的样本数量要求较低,并且极大地减少了训练时间,拓宽了深度模型的应用领域。另外,由

于决策树构成了深度森林的基本单元,使得该模型具有较好的可解释性,需要人工设置的超参数数量也非常少,在许多缺少历史经验的领域也取得了较好的效果。目前,深度森林模型已经在医疗、交通、金融等多个领域得到了应用<sup>[13-15]</sup>。

## 2 电力负荷预测模型

### 2.1 多粒度级联森林

文中基于多粒度级联森林(gcForest, multi-Grained Cascade Forest)模型构建电力负荷预测模型。gcForest 主要由两个结构组成:多粒度扫描窗口和级联森林。多粒度扫描窗口类似于卷积神经网络中的滑动窗口,对输入的数据进行不同粒度的采样,获取输入特征,增强学习效果。级联森林的作用是对输入特征进行表征学习,它由多种随机森林构成。如图 1 所示,每一层都包括 2 个随机森林和 2 个完全随机森林,每个随机森林包含若干棵树,而每个森林都会输出一个类向量结果。将多个森林得到的类向量进行串联,即为这一层的输出。将其与最初的输入特征连接作为下一层的输入,这样,每一层都会接收到前一层的特征信息和最初的特征信息。模型会一直加深层数直到收敛或者到达指定的层数上限为止,最后一层的输出类向量就是分类的最终结果。

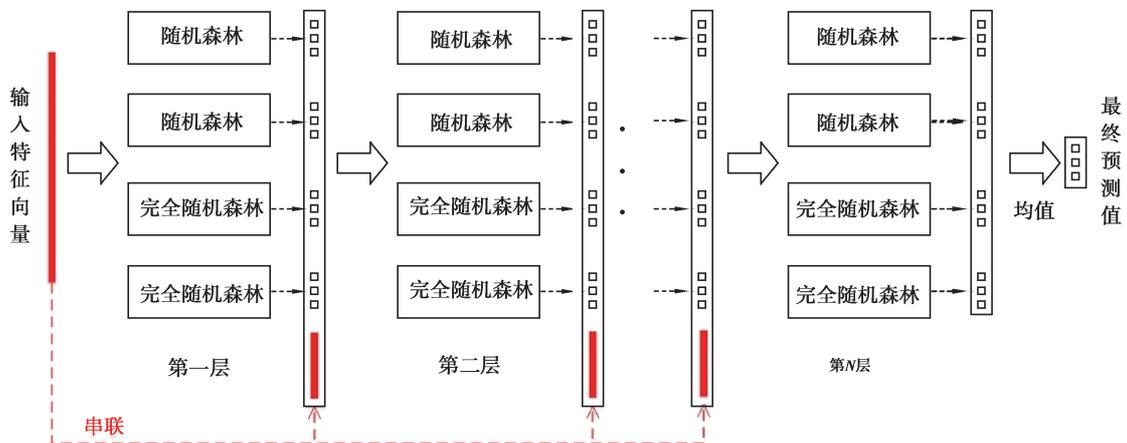


图 1 级联森林结构图

Fig. 1 Structure of the cascade forest

### 2.2 改进的深度森林算法的负荷预测模型

根据上述介绍可知,深度森林的输入特征依赖于多粒度扫描窗口的设计。不同于图像识别领域,负荷预测模型的输入通常是一个一维向量,因此需要根据负荷数据的特点来设计有效的多粒度扫描窗口。

电力负荷预测具有周期性和时变性的特点,因此,负荷预测需要以历史数据和客观因素为参考开展预测,需要在预测前就明确采用历史数据的时间段和未来预测的时间范围。由于负荷数据是具有时间连续性的一维向量,选择合适的窗口才能有效利用历史数据的周期性。在文中,历史数据为预测节点前 18 个月的电力负荷值,预测目标设定为短期预测,未来预测时间从一小时到一周不等。

针对电力负荷数据的特性,对深度森林算法进行改进。原 gcForest 模型通过调整窗口的大小来提取数据特征,而文中将在改变窗口大小的同时,通过调整窗口滑动的步长来提取数据的周期性。由于电力负荷数据的周期性与人类的生产生活规律密切相关,在天、周、月、年等不同的时间跨度中都表现出一定的周期性。为了提取不同粒度的周期性,参照人类对于时间的划分方式,设定不同的步长,包括一小时、一天、一周、一月、一季以及一年,并对不同的滑动步长设定对应的窗口大小。

由于输入向量的长度较长, $365 \text{ d} \times 24 \text{ h} \times 4 = 35\ 040 \text{ h}$ 。当扫描步长较小时,产生的特征将特别巨大。以三分类为例,以步长 4(一小时)为例,设窗口大小为 96,则最终生成的输入特征为 $((35\ 040 - 96) \div 4 + 1) \times 3 = 26\ 211$  维。若将其作为级联森林的输入特征,会极大地增加时间复杂度和空间复杂度,甚至出现内存溢出的情况。针对这一问题,文中通过设定滑动次数上限的方式来解决。文中的窗口大小、步长以及滑动次数

上限设定如表 1 所示。

表 1 多粒度窗口滑动步长与滑动次数

Table 1 The sliding step length and sliding times of multi-granularity window

窗口大小	步长	滑动次数上限
24(6 小时)	4	672
96(1 天)	24	120
672(1 周)	672	48
2 880(1 月)	2 880	无

如图 2 所示,以窗口大小为 24 和 672 为例,分别得到  $1\ 344\ m$  维和  $96\ m$  维的特征,其中  $m$  是类向量长度,即分类的类别数量。串联组合得到级联森林部分的输入特征。

此外,若是直接以类别中间值为最后的预测结果,则会人为造成大量的误差。针对这一问题,文章改进了最后一层的输出。深度森林中每一个随机森林的输出是根据其中每一棵树得到的类向量求均值,将离散的模型输出变为连续值,进而能够提供更为精确的预测结果。文中最后一层的树不会输出类向量,而是通过计算对应叶子节点中的具体数值来得到最终的预测值。如图 3 所示,其中红色线段代表测试样本落入叶子节点的路径。在多粒度窗口扫描的过程中,按照类向量的生成方法得到最初的数据特征。在预测过程中,按照均值的方法得到最终的预测值。

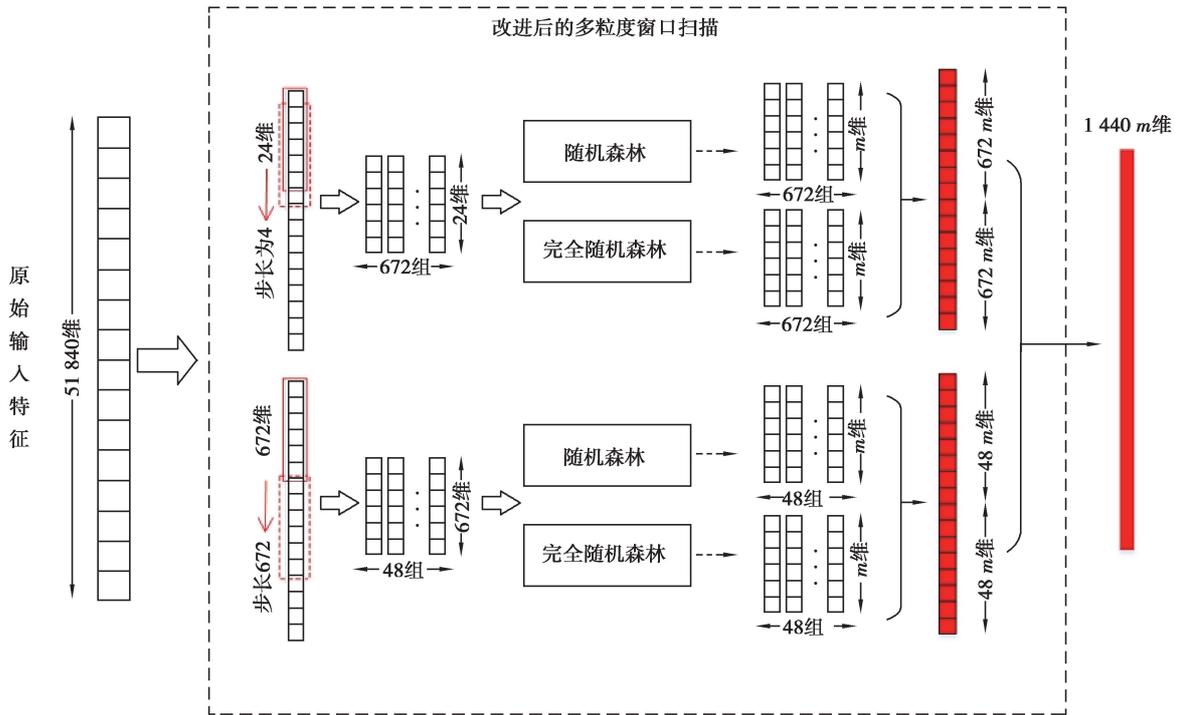


图 2 多粒度滑动窗口扫描方法

Fig. 2 Illustration of multi-granularity sliding window scan

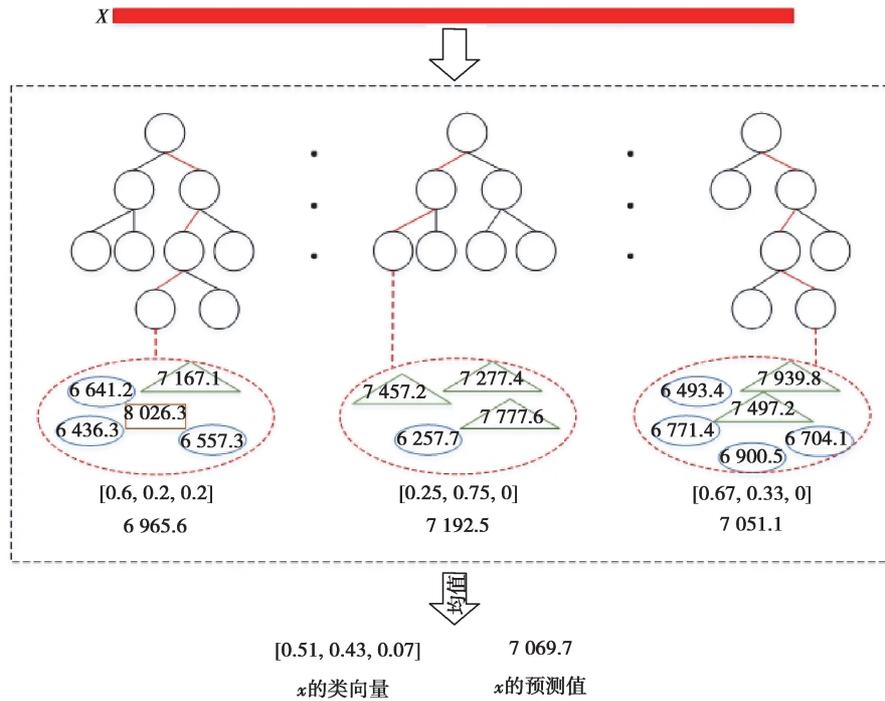


图 3 类向量生成方法及改进后的预测值计算方法

Fig. 3 Illustration of class vector generation and predicted value calculation

### 3 实 验

#### 3.1 实验数据

文中数据来源于国家电网公司东北分部,包括了从 2016 年到 2019 年上半年的电力负荷数据,涵盖了辽宁、吉林、黑龙江和内蒙古东部地区。数据具有较强的时间连续性,每过 15 min 记录一次当前电网的功率值,总计应该有 612 960 条数据,实际测得 567 168 条数据,数据较为完整、质量较高,能够用于验证模型的可行性。

##### 3.1.1 数据预处理

由于传感器故障和异常工况等原因,原始数据包含一些错误信息和数据缺失的情况,必须对原始数据进行数据清洗以减小错漏数据的影响。由于历史数据具有时间连续性,对错误数据并不能单纯地剔除处理,因此文中针对错漏数据采用线性插值的方法进行数据填充。在处理数据时,没有出现连续大段的错漏数据,因此用插值法是可行的。

##### 3.1.2 数据分段

为方便构建决策树,减少计算量,通过将不同的负荷状态进行分段,将负荷预测问题转化为分类问题。例如,以功率 1 000 为区间单位长度,对于辽宁省则分为 10 000~30 000 共计 31 个区间,也就是 31 类;黑龙江省为 1 000~20 000 共 21 个类别;内蒙东部和吉林分为 1 000~15 000 共计 16 个类别;总负荷分为 20 000~60 000 共计 41 个类别。分类区间越短,预测精度越高,但准确率也会下降。

#### 3.2 实验设计

实验将分别对 4 个地区的电力负荷数据与东北电网总负荷数据集进行验证,预测目标为 2019 年上半年的电力负荷数据,输入数据包括 2018 年全年的电力负荷数据,预测时间设为一天。对于改进后的深度森林算法,若预测值与真实值的误差绝对值小于分段区间的一半,则认为预测结果正确,通过这种方法,可以统一以分类的准确率为标准进行比较。

对照实验设计如下:

1)k 近邻方法(KNN, k-nearest neighbor):该方法是一种简单有效的分类聚类方法,能够有效应用于具有大量训练集的任务中并能够取得不错的结果,在许多领域都有应用,因此文中选择该方法作为传统机器学习的典型算法之一。

2)支持向量机(SVM):支持向量机作为最常用的分类模型,在许多领域都验证了它的有效性和可靠性,其中的参数都使用交叉验证的方法调至最优。通过与该方法进行对比可以验证文中所提出方法的有效性。

3)随机森林(RF, random forest):深度森林的基础模型,对于一些较为简单的问题往往能取得较好的分类效果。实验中随机森林的参数设定与深度森林中单一森林的参数设定相同,可以验证深度森林这一集成思想的有效性。

4)栈式自动编码器(SAE, stacked autoencoder):该方法是一种典型的深度特征学习方法,能够有效提取数据特征。与该方法进行对比,可以验证文中所提方法相对于传统深度学习模型的优势。

5)多粒度级联森林(gcForest):针对电力负荷数据特点对改进后的深度森林模型进行有效性验证,将所提出方法与基本的 gcForest 模型进行对比。参数设定与文中提出方法一致,只是窗口大小固定为 192,步长固定为 48。

### 3.3 参数设置

文章对改进的深度森林方法(DF)主要采用默认的超参数设置,具体参数设置如表 2 所示。

表 2 改进的深度森林方法的超参数设置  
Table 2 Hyper-parameters setting of improved deep forest

阶段	超参数名称	参数值
多粒度扫描	森林数量	2
	决策树数量	30
	决策树深度上限	100
	滑动窗口设置	见表 1
级联森林	森林数量	4
	决策树数量	500
	决策树深度上限	无

### 3.4 实验结果

应用深度森林算法对东北总电网的功率进行预测,并于多种机器学习方法进行对比,结果如表 3 所示。在分段区间大小为 9 000 时,改进的深度森林方法取得了较好的预测结果。在缩小分段区间大小即提高预测精度要求的过程中,对比实验的准确率会快速下降,而文章所提出的方法却仍然能够维持较高的准确率,说明改进的深度森林方法能够有效提升电力负荷预测的精度,验证了模型的有效性。

表 3 东北总电网负荷预测结果  
Table 3 Load forecast results of northeast power grid

模型方法	精度 9 000	精度 6 000	精度 3 000
KNN	0.775	0.582	0.430
SVM	0.758	0.643	0.547
RF	0.774	0.693	0.382
SAE	0.785	0.742	0.503
gcForest	0.874	0.743	0.563
DF	0.864	0.764	0.648

图 4 为其他地区的电网数据上的实验结果。从图中可以更加明显地看出,当预测精度要求不高时,传统的机器学习方法与深度学习方法都能取得较高的准确率,但是当逐渐提高精度要求,传统学习方法的准确率会快速下降,难以取得较为准确的预测结果。深度学习方法较传统学习方法能够获得更好的特征,进而能够满足更高的精度要求。而改进的深度森林算法在分段区间较小时,能够取得较为明显的优势。当预测精度

范围小于波动范围 20%时,传统的深度学习方法的准确率已经出现明显下降,而文中所提出的方法在波动范围 10%的精度要求下仍能维持较高的准确率。

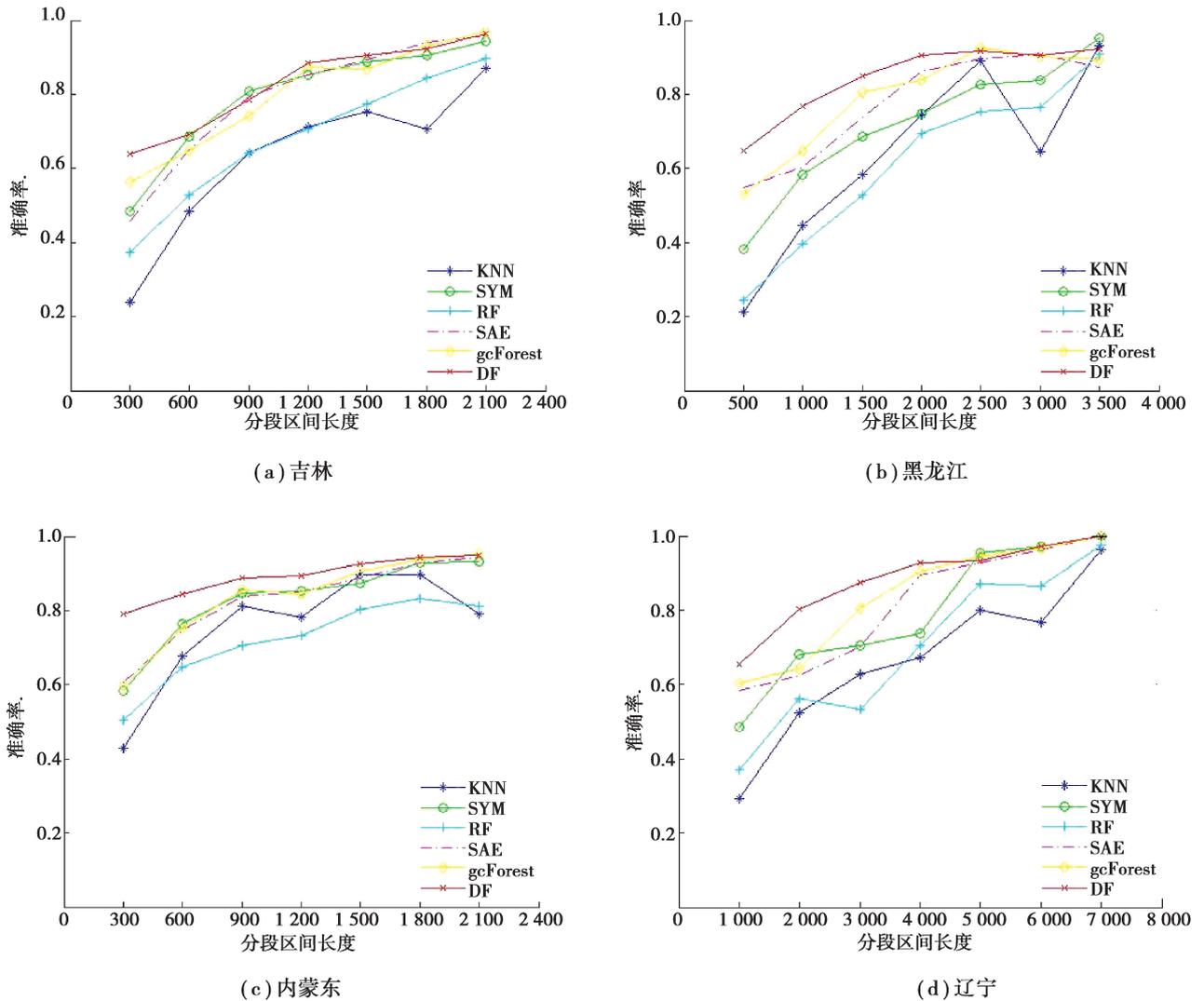


图 4 各地区电网负荷预测结果

Fig. 4 The accuracy of load forecasting

表 4 为深度算法模型的层数。相比于传统的深度学习方法,深度森林由于不需要预设模型结构,减少了超参数的设定,训练得到的模型能够更加适合数据集,提升算法的效率与准确度。而文中改进了深度森林的多粒度窗口扫描方法,利用电力负荷预测方面的背景知识,在模型的第一层消除了大量信息密度较低的数据,进一步提升了模型运行的效率,加快模型收敛速度,减少时间开销的同时也降低了对硬件的需求,在此验证了文中所提出的改进的深度森林方法在电力负荷预测方面的有效性。

表 4 深度模型层数对比

Table 4 The number of layers in the deep models

数据集	SAE	gcForest	DF
总负荷	16	17	13
吉林电网	16	9	6
黑龙江电网	16	9	7
辽宁电网	16	13	7
内蒙古电网	16	14	9

## 4 结束语

文章进行了电力负荷预测技术的研究,针对深度学习算法参数多、收敛慢、数据量要求大的问题,将深度森林方法引入了电力短期负荷预测领域,并针对 gcForest 模型在电力负荷预测领域的不足对模型进行改进。选取中国东北三省以及内蒙古东部的实际电网数据进行验证,并与传统机器学习模型进行了对比,通过实验验证了所提出的负荷预测算法在短期电力预测方面的优异性。

### 参考文献:

- [1] 杨博宇, 陈仕军. 电力负荷预测研究综述及预测分析[J]. 四川电力技术, 2018, 41(3): 56-60, 91.  
Yang B Y, Chen S J. Overview of electrical load forecasting researches and forecasting analysis[J]. Sichuan Electric Power Technology, 2018, 41(3): 56-60, 91. (in Chinese)
- [2] 牟丰. 基于机器学习的短期电力负荷预测算法研究[D]. 西安:西安科技大学, 2018.  
Mou F. Research on short-term power load forecasting algorithm based on machine learning[D]. Xi'an: Xi'an University of Science and Technology, 2018. (in Chinese)
- [3] 孙建梅, 钱秀婷, 王永晴. 基于改进灰色模型的中长期电力负荷预测[J]. 电工技术, 2019(19): 28-31.  
Sun J M, Qian X T, Wang Y Q. Medium and long term power load forecasting based on improved grey model[J]. Electric Engineering, 2019(19): 28-31. (in Chinese)
- [4] 张志, 延菱, 崔慧军, 等. 考虑关联因素的智能化中长期电力负荷预测方法[J]. 电力系统保护与控制, 2019, 47(2): 24-30.  
Zhang Z, Du Y L, Cui H J, et al. Intelligent mid-long electricity load forecast method considering associated factors [J]. Power System Protection and Control, 2019, 47(2): 24-30. (in Chinese)
- [5] 邢海军, 程浩忠, 张沈习, 等. 主动配电网规划研究综述[J]. 电网技术, 2015, 39(10): 2705-2711.  
Xing H J, Cheng H Z, Zhang S X, et al. Review of active distribution network planning[J]. Power System Technology, 2015, 39(10): 2705-2711. (in Chinese)
- [6] 罗宁, 高华, 贺墨琳. 基于神经网络的电力负荷预测方法研究[J]. 自动化与仪器仪表, 2020(1): 157-160.  
Luo N, Gao H, He M L. Research on power load forecasting method based on neural network[J]. Automation & Instrumentation, 2020(1): 157-160. (in Chinese)
- [7] 简献忠, 顾祎婷. 一种基于全过程优化支持向量机的短期电力负荷预测方法[J]. 电力科学与工程, 2018, 34(11): 45-51.  
Jian X Z, Gu Y T. Short-term electric load forecasting method based on global optimization support vector machine [J]. Electric Power Science and Engineering, 2018, 34(11): 45-51. (in Chinese)
- [8] 张静, 石鑫. 基于改进 MOPSO-BP 算法的短期电力负荷预测研究[J]. 电力学报, 2019, 34(6): 556-563.  
Zhang J, Shi X. Short-term power load forecasting based on improved MOPSO-BP algorithm[J]. Journal of Electric Power, 2019, 34(6): 556-563. (in Chinese)
- [9] Kong W, Dong Z Y, Jia Y, et al. Short-term residential load forecasting based on LSTM recurrent neural network [J]. IEEE Transactions on Smart Grid, 2017, 10(1): 841-851.
- [10] Kuo P H, Huang C J. A high precision artificial neural networks model for short-term energy load forecasting[J]. Energies, 2018, 11(1): 213.
- [11] Chen K, Chen K, Wang Q, et al. Short-term load forecasting with deep residual networks[J]. IEEE Transactions on Smart Grid, 2018, 10(4): 3943-3952.
- [12] Zhou Z H, Feng J. Deep forest[J]. arXiv preprint arXiv:1702.08835, 2017.
- [13] Guo Y, Liu S, Li Z, et al. BCDForest: a boosting cascade deep forest model towards the classification of cancer subtypes based on gene expression data[J]. BMC bioinformatics, 2018, 19(5): 118.
- [14] Su R, Liu X, Wei L, et al. Deep-resp-forest: a deep forest model to predict anti-cancer drug response[J]. Methods, 2019, 166: 91-102.
- [15] Tanno R, Arulkumaran K, Alexander D C, et al. Adaptive neural trees[J]. arXiv preprint arXiv:1807.06699, 2018.