文章编号：1006－9798(2019)03－0007－04；DOI：10.13306/j.1006－9798.2019.03.002

基于改进随机森林算法的电力系统短期负荷预测模型

邢书豪[1]，孙文慧[2]，颜　勇[3]，张智晟[1]
(1. 青岛大学电气工程学院，山东 青岛 266071；2. 青岛地铁集团有限公司，山东 青岛 266000；
3. 国网山东综合能源服务有限公司，山东 济南 250000)

摘要：为了提高电力系统短期负荷预测的准确性，本文提出了基于改进随机森林算法的电力系统短期负荷预测模型。改进随机森林算法是将随机森林算法中的决策树数量和分裂特征数等参数采用粒子群进行优化，通过比较每组参数对应的随机森林袋外数据误差，获取参数最优值，使随机森林算法的性能得到最优，并采用山东省某城市电网的历史负荷数据进行仿真分析。仿真结果表明，与基于传统随机森林算法的预测模型相比，本文所提出的预测模型的平均绝对误差降低0.81%，最大相对误差降低1.89%，说明本文所提出的基于改进随机森林算法的短期负荷预测模型具有更好的预测性能。该研究具有一定的工程实用性。

关键词：改进随机森林算法；粒子群优化算法；短期负荷预测；电力系统
中图分类号：TM715　　**文献标识码**：A

电力系统短期负荷预测对电力系统稳定运行具有至关重要的影响。准确的短期负荷预测，不仅是电力市场发电计划制订的基础，而且还可保障电力系统的安全运行[1]，一直是电力系统的重要研究内容[2]。近年来，对于电力系统短期负荷预测，国内外专家学者提出了诸多预测模型，包括人工神经网络[3-4]、支持向量机[5-6]、小波分析法[7]等。随机森林算法是基于传统决策树的统计学习理论，是一种新兴智能算法，它可有效处理高维数据，鉴于它具有较高的准确率，克服了过拟合的问题，目前已被广泛应用于医学、经济学、水文科学、生物信息等领域[8-10]。同时，它在短期负荷预测的研究中也具有较好的预测效果[11]。但是针对不同的研究负荷对象，算法中的决策树数量和分裂特征数等参数对模型性能影响较大[12]，而传统随机森林算法存在根据经验选取决策树数量和分裂特征数等参数问题，进而导致随机森林算法达不到性能最优。因此，本文提出了基于改进随机森林算法的电力系统短期负荷预测模型，随机森林算法中的决策树数量和分裂特征数等参数采用粒子群优化算法进行优化，获取参数最优值，使随机森林算法的性能得到最优。通过算例仿真表明，与传统预测模型相比，本文所提出的该模型取得了较为满意的预测效果。该研究对电力系统稳定运行具有重要意义。

1　随机森林算法

2001年，随机森林算法(random forest，RF)由L.Breiman等人[13]提出。该算法是一种基于传统决策树的统计学习理论，具有较强的随机性，其随机性主要体现在以下两个方面[14]：一是从原始样本集中采用有放回的方式，随机选取样本数据构成训练样本集；二是候选分裂属性，由生成决策树时随机选取的特征属性产生。

在算法中，随机森林实际上是一个分类器集合，它由众多决策树分类器 $h(x,\theta_k)$，$k=1,2,\cdots,n$ 组成，其中 θ_k 表示独立且相同分布的随机变量，每个决策树分类器都对输入变量 x 的类别归属进行预测。随机森林通过Bagging方法，生成彼此之间互不相同的训练样本集，分类回归树作为元分类器组合为集成分类器，预测结果由所有分类器求算数平均值所得[15]。

1.1　Bagging方法取样

Bagging(bootstrap aggregating)方法是一种有放回抽样方法，即以可重复的随机抽样为基础，每个样本都由初始数据集进行有放回抽样得到。该方法是采用Bootstrap方法，从原始样本集中随机抽选n个训练样本，并将

收稿日期：2018－10－12；修回日期：2019－02－15
基金项目：国家自然科学基金资助项目(51477078)；智能电网教育部重点实验室开放研究基金(2018)
作者简介：邢书豪(1994－)，男，山东滨州人，硕士研究生，主要研究方向为电力系统短期负荷预测。
通信作者：张智晟(1975－)，男，山东青岛人，博士，教授，主要研究方向为电力系统短期负荷预测和调度运行。Email：slnzzs@126.com

该过程进行 k 次循环,从而得到 k 个训练集[16]。在生成训练子集时,每个训练样本都有可能被抽取,但是当多次重复训练时,将总会有一部分样本未被抽取,样本不被抽取的概率为 $(1-(1/n))^n$,其中 n 是初始数据中的样本总数。Bootstrap 重抽样示意图如图 1 所示。

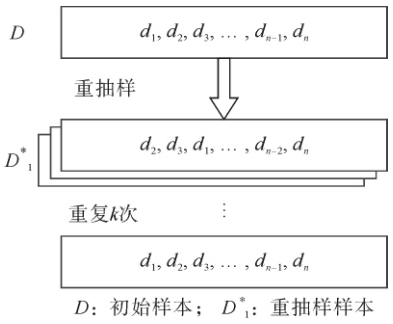

图 1 Bootstrap 重抽样示意图

1.2 CART 决策树

分类回归树(classification and regression trees,CART)算法是利用二分递归分割方法,把原样本集划分为 2 个子集,从而会有 2 个分支在每个非叶子节点上面。在节点分裂的时候,分裂规则按照 Gini 指标最小原则,概率分布的 Gini 指数可计算为

$$\mathrm{Gini}(p)=\sum_{k=1}^{K}p_k(1-p_k)=1-\sum_{k=1}^{K}p_k^2 \tag{1}$$

式中,K 为节点中特征样本的总种类数;p_k 为属于节点中第 k 类特征样本的概率。样本集合 D 的 Gini 指数可计算为

$$\mathrm{Gini}(D)=1-\sum_{k=1}^{K}\left(\frac{|C_k|}{|D|}\right)^2 \tag{2}$$

式中,C_k 为样本集合 D 中属于第 k 类的样本子集。每个划分的 Gini 指数可计算为

$$\mathrm{Gini}_{\mathrm{split}}(D)=\frac{|D_1|}{|D|}\mathrm{Gini}(D_1)+\frac{|D_2|}{|D|}\mathrm{Gini}(D_2) \tag{3}$$

式中,D_1 和 D_2 为样本集合 D 分割成的 2 个子集。

1.3 随机森林的构成

设随机森林由一系列 CART 树 $h(x,\theta_k)$,$k=1,2,\cdots,n$ 构成,其边缘函数可表征为

$$K(\boldsymbol{X},\boldsymbol{Y})=av_kI(h(\boldsymbol{X},\theta_k)=\boldsymbol{Y})-\max_{j\neq Y}av_kI(h(\boldsymbol{X},\theta_k)=j) \tag{4}$$

其中,$\boldsymbol{X}$ 为输入向量,最多包含 J 种不同的类别;j 为 J 种类别中的某一类;$\boldsymbol{Y}$ 为正确的分类向量;$I(\cdot)$为指示函数;$av_k(\cdot)$为取平均函数。随机森林的泛化误差可表征为

$$PE^{\cdot}=P_{X,Y}(K(\boldsymbol{X},\boldsymbol{Y})<0) \tag{5}$$

其中,$P_{X,Y}$为对给定输入变量 $\boldsymbol{X}$ 的分类错误率函数。随机森林的泛化误差最大值可表征为

$$PE^{\cdot}\leqslant\frac{\bar{\rho}(1-s^2)}{s^2} \tag{6}$$

其中,$\bar{\rho}$ 为决策树平均相关系数;s 为决策树的平均强度。

式(6)表明,随机森林的泛化误差最大值与决策树的平均相关系数以及决策树的平均强度有关,也就是说当决策树的平均相关系数越小、决策树的平均强度越高时,随机森林的泛化性能越好。因此,可通过决策树平均相关系数的降低和决策树平均强度的提升,来实现随机森林预测精度的提高。

2 基于改进随机森林算法的电力系统短期负荷预测模型

本文提出了基于改进随机森林算法的电力系统短期负荷预测模型,模型中随机森林算法中的决策树数量和分裂特征数采用粒子群优化算法进行获取,使模型预测性能最优。

粒子群优化算法(particle swarm optimization,PSO)是一种模拟鸟群飞行捕食行为的优化算法[17]。该算法把整个粒子群比作鸟群,种群中的每一个粒子都代表解可行域中的一个可行解,但不一定是最优解。粒子在每次循环运算时,通过学习自身历史经验和种群历史经验,与上一位置的适应值作比较,从而调整自身的速度和位置矢量,最终达到全局寻优的目的。

假设种群内存在的粒子数为 L,则 M 维解空间中的第 i 个粒子位置矢量可以定义为 $\boldsymbol{x}_i=(x_{i1},x_{i2},\cdots,x_{id})$,其速度为 $\boldsymbol{v}_i=(v_{i1},v_{i2},\cdots,v_{id})$,其中 $i=(1,2,\cdots,L)$,$d=(1,2,\cdots,M)$。将位置矢量 $\boldsymbol{x}_i$ 代入目标函数,可完成对解的优劣评价,第 i 个粒子通过循环迭代得到的最优解 $\boldsymbol{P}_i=(P_{i1},P_{i2},\cdots,P_{id})$,即该粒子的个体极值 p_{best};整个粒子群搜索到的最佳位置 $\boldsymbol{P}_g=(P_{g1},P_{g2},\cdots,P_{gd})$,即该种群的全局极值 g_{best}。

在循环迭代过程中,粒子的速度和位置矢量更新为

$$\boldsymbol{v}_{id}=\omega*\boldsymbol{v}_{id}+c_1*\mathrm{rand}()*(p_{id}-\boldsymbol{x}_{id})+c_2*\mathrm{rand}()*(p_{gd}-\boldsymbol{x}_{id}) \tag{7}$$

$$\boldsymbol{x}_{id}=\boldsymbol{x}_{id}+\boldsymbol{v}_{id} \tag{8}$$

式中，ω 为惯性权重，对种群中全局搜索和局部细化能力起到关键作用；c_1 和 c_2 为加速常数，也称作学习因子；rand()为在[0,1]区间内波动的随机函数[18]。

本文针对随机森林算法中决策树数量 k 和分裂特征数 m 为离散值的特点，采用粒子群优化算法进行参数优化。通过比较对应于每次参数袋外数据误差大小，检验预测性能的优劣。袋外数据误差映射了随机森林算法的预测性能，该袋外数据误差愈小，随机森林算法预测性能愈佳[19]。袋外数据误差为

$$e_{\mathrm{OOB}}=\sum_{i}^{k}e_{\mathrm{OOB}}(i)/k \tag{9}$$

其中，$e_{\mathrm{OOB}}(i)$为第 i 棵决策树的袋外数据误差。改进随机森林算法流程如图 2 所示。

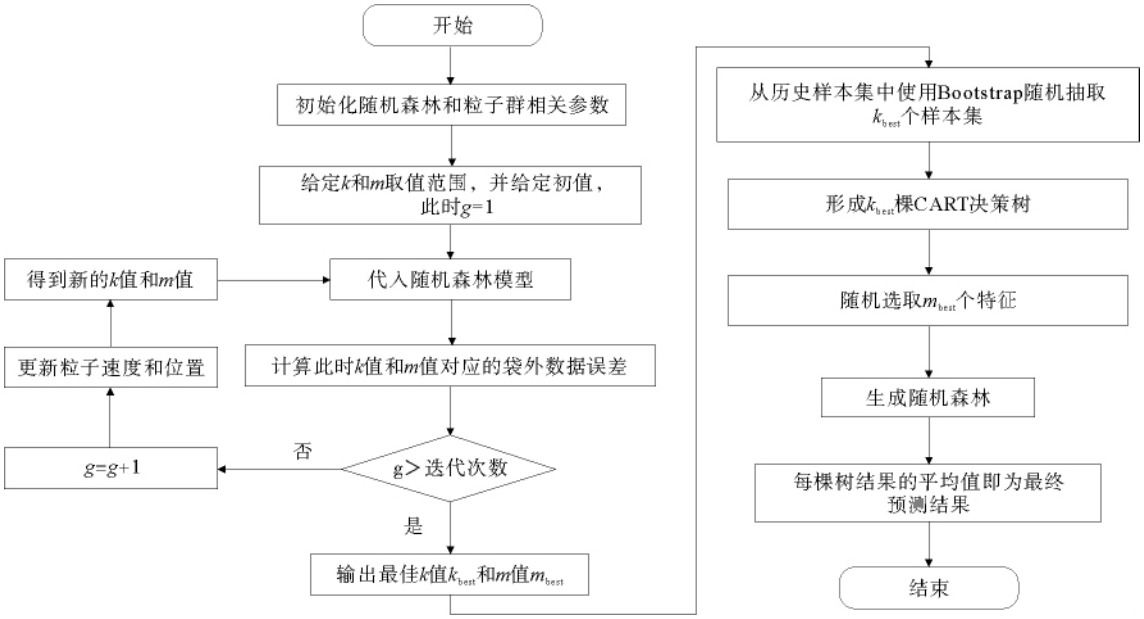

图 2　改进随机森林算法流程图

3　算例分析

3.1　数据集及数据处理

本文采用山东省某城市电网的历史负荷数据，输入数据包含每日 96 点负荷值及日最高温度、日最低温度、日平均温度、日降水概率和日类型等 5 类外因数据。因各类数据量纲存在不同，故需将所有输入数据采用归一化处理，使处理后各类数据的数值在[0,1]之间，负荷数据和温度数据归一化可表征为

$$l=\frac{l_{\mathrm{in}}-l_{\mathrm{min}}}{l_{\mathrm{max}}-l_{\mathrm{min}}} \tag{10}$$

其中，l 为归一化处理之后的输入数据；l_{in}为原始输入数据；l_{max}为原始输入数据中的最大值；l_{min}为原始输入数据中的最小值[20]。日降水概率取值区间为[0,1]。日类型分为工作日(星期一到星期五)取 1，休息日(星期六、星期日)取 0.5。

3.2　算例结果分析

本文选取待预测日前 3 天的数据来预测待预测日 1 天 96 个点的负荷数据，为预测待预测日 t 时刻的负荷数据，输入变量为 29 维，即 $\boldsymbol{X}=[x_1\quad x_2\quad x_3\cdots x_{29}]$，其中 $x_1\sim x_9$ 表示待预测日前 3 天中的 $t-1$，t，$t+1$ 这 3 个时刻的历史负荷数据；$x_{10}\sim x_{18}$表示待预测日前 3 天的日最高温度、日最低温度和日平均温度；$x_{19}\sim x_{21}$表示待预测日前 3 天的日降水概率；$x_{22}\sim x_{24}$表示待预测日前 3 天的日类型；$x_{25}\sim x_{29}$表示预测日的日最高温度、日最低温度、日平均温度、日降水概率和日类型。本文算例中，预测日当天日最高温度为 17 ℃，日最低温度为 9 ℃，日平均温度为 13 ℃，日降水概率为 0.6，日类型为 1。

为研究本文所采用模型的预测效果,将基于传统随机森林算法的短期负荷预测模型(模型1)和本文提出的基于改进随机森林算法的短期负荷预测模型(模型2)的预测效果进行对比。两种预测模型预测值与实际值曲线对比如图3所示。图3中,模型1预测的负荷平均绝对误差为2.66%,最大相对误差为6.63%;模型2预测的负荷平均绝对误差为1.85%,最大相对误差为4.74%。与模型1相比,模型2预测的负荷平均绝对误差降低0.81%,最大相对误差降低1.89%,可见本文所提出的基于改进随机森林算法的短期负荷预测模型具有更好的预测性能。

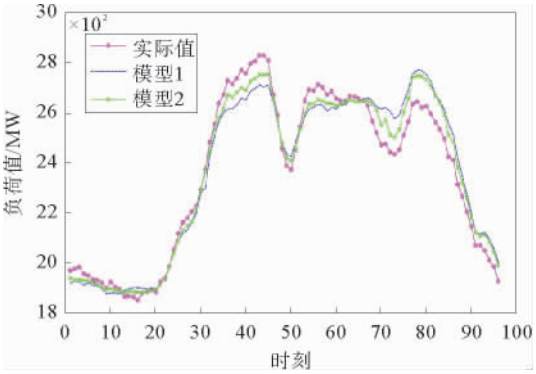

图3 两种预测模型预测值与实际值曲线对比

4 结束语

本文提出了基于改进随机森林算法的电力系统短期负荷预测模型。该模型采用粒子群优化算法对随机森林中不同的决策树数量和分裂特征数等相关参数进行优化,弥补了根据经验选取随机森林相关参数而导致模型性能不佳的不足,进一步提高了模型的预测性能。通过与基于传统随机森林算法的预测模型对比,验证了本文所提出的模型具有更好的预测精度。

参考文献:

[1] 康重庆,夏清,刘梅. 电力系统负荷预测[M]. 北京:中国电力出版社,2017.
[2] 廖旎焕,胡智宏,马莹莹,等. 电力系统短期负荷预测方法综述[J]. 电力系统保护与控制,2011,39(1):147-152.
[3] 陈夫进,王宝成. 基于BP神经网络系统的短期电力负荷预测[J]. 河南科学,2013,31(2):168-171.
[4] 张亚军,张大波,许诚昕. 神经网络在电力系统短期负荷预测中的应用综述[J]. 浙江电力,2007,26(2):5-9.
[5] 何习佳. 基于支持向量机的电力短期负荷预测研究[J]. 电子设计工程,2009,17(12):90-92.
[6] 王小君,毕圣,徐云鹃,等. 基于数据挖掘技术和支持向量机的短期负荷预测[J]. 电测与仪表,2016,53(10):62-67.
[7] 李博,门德月,严亚勤,等. 基于数值天气预报的母线负荷预测[J]. 电力系统自动化,2015,39(1):137-140.
[8] 方匡南,吴见彬,朱建平,等. 随机森林方法研究综述[J]. 统计与信息论坛,2011,26(3):32-38.
[9] 余坤勇,姚雄,邱祈荣,等. 基于随机森林模型的山体滑坡空间预测研究[J]. 农业机械学报,2016,47(10):338-345.
[10] 巩亚楠,帕提麦·马秉成,朱登浩,等. 随机森林与Logistic回归在预约挂号失约影响因素预测中的应用[J]. 现代预防医学,2014,41(5):769-772.
[11] 李婉华,陈宏,郭昆,等. 基于随机森林算法的用电负荷预测研究[J]. 计算机工程与应用,2016,52(23):236-243.
[12] 温博文,董文瀚,解武杰,等. 基于改进网格搜索算法的随机森林参数优化[J]. 计算机工程与应用,2018,54(10):154-157.
[13] Breiman L. Random Forests[J]. Machine Learning,2001,45(1):5-32.
[14] 董师师,黄哲学. 随机森林理论浅析[J]. 集成技术,2013,2(1):1-7.
[15] 吴潇雨,和敬涵,张沛,等. 基于灰色投影改进随机森林算法的电力系统短期负荷预测[J]. 电力系统自动化,2015,39(12):50-55.
[16] Breiman L. Bagging predictors[J]. Machine Learning,1996,24(2):123-140.
[17] 李爱国,覃征,鲍复民,等. 粒子群优化算法[J]. 计算机工程与应用,2002,38(21):1-3,17.
[18] 程学新. 粒子群优化加权随机森林算法研究[D]. 郑州:郑州大学,2017.
[19] 马骊. 随机森林算法的优化改进研究[D]. 广州:暨南大学,2016.
[20] 于惠鸣,撖奥洋,于立涛,等. 基于PSO-DNN的电力系统短期负荷预测模型研究[J]. 青岛大学学报:工程技术版,2017,32(2):62-66.

(下转第38页)

电子文献载体和标志代码

磁带(magnetic tape)	磁盘(disk)	光盘(CD-ROM)	联机网络(online)
MT	DK	CD	OL

Design of Rail Traffic Data Monitoring and Processing System Based on LabVIEW

CHONG Xingjing[1], LENG Ziwen[2], GAO Junwei[1], GAI Hongyu[1]
(1. School of Automation, Qingdao University, Qingdao 266071, China);
(2. Rizhao Branch of Shandong Special Equipment Inspection Institute, Rizhao 276800, China)

Abstract: In order to facilitate the maintenance of rail transit train monitoring and auxiliary inverters, this paper designs a real-time display of rail transit train auxiliary inverter monitoring system based on LabVIEW. The computer and NI cRIO 9075 data acquisition card are used to build the hardware system. The LabVIEW development platform is used to establish a visual human-machine interface, and the simulation system is simulated in the intelligent system laboratory in combination with the database management system. The simulation results show that the monitoring system can accurately display the physical quantity such as ambient temperature and current, realize the display and storage of data, monitor the change of auxiliary inverter voltage in real time, analyze the data from time domain and frequency domain, and extract useful signals. This study laid the theoretical foundation for further fault signal diagnosis and analysis.

Key words: LabVIEW; auxiliary inverter; data acquisition; monitoring system

---

(上接第10页)

Short-Term Load Forecasting Model of Power System Based on Improved Random Forest Algorithm

XING Shuhao[1], SUN Wenhui[2], YAN Yong[3], ZHANG Zhisheng[1]
(1. College of Electrical Engineering, Qingdao University, Qingdao 266071, China;
2. Qingdao Metro Group Co., Ltd., Qingdao 266000, China;
3. State Grid Shandong Integrated Energy Service Co., Ltd., Jinan 255000, China)

Abstract: In order to improve the accuracy of power system short-term load forecasting, this paper proposes a short-term load forecasting model based on improved random forest algorithm. The improved random forest algorithm is to optimize the parameters such as the number of decision trees and the number of splitting features in the random forest algorithm by using particle swarm optimization. By comparing the errors of the data outside the bag corresponding to each group of parameters, the optimal values of parameters are obtained, so that the performance of the random forest algorithm is optimized, and the historical load data of a city power grid in Shandong Province is used for simulation analysis. The simulation results show that the average absolute error and the maximum relative error of the forecasting model proposed in this paper are reduced by 0.81% and 1.89% respectively, compared with the traditional random forest algorithm, which shows that the short-term load forecasting model based on the improved random forest algorithm proposed in this paper has better forecasting performance. The model has certain engineering practicability.

Key words: improved random forest algorithm; particle swarm optimization; short-term load forecasting; power system