

基于配电物联网技术和并行随机森林算法的 日前V2G功率容量预测

岳友¹, 李宽宏²

(1. 南瑞集团(国网电力科学研究院)有限公司, 南京 211100;

2. 国网福建省电力有限公司 福州供电公司, 福州 350000)

Day-ahead prediction of V2G power capacity based on distribution Internet of Things technology and parallel random forest algorithm

YUE You¹, LI Kuanhong²

(1. NARI Group Corporation/State Grid Electric Power Research Institute, Nanjing, 211100, China;

2. State Grid Fujian Electric Power Co., Ltd., Fuzhou Electric Power Supply Company, Fuzhou, 350000, China)

摘要:实现大量电动汽车所聚合电池储能作为电网和可再生能源的缓冲,关键是精确和快速的V2G功率容量预测。然而,电动汽车充放电行为的随机性和不确定性,给精确的V2G功率预测带来了困难。随着配电物联网建设的推进和电动汽车的大规模接入,电动汽车充放电数据爆发式增长,为精度的V2G功率容量提供大量数据支撑的同时,也带来了大数据处理问题。通过利用大量历史和气象数据构建预测特征向量,基于并行随机森林算法建立了日前V2G功率容量的并行预测模型。从而,避免了用户行为对预测结果的影响。此外,基于Spark搭建了分布式的大数据平台,以实现V2G功率容量快速预测。最终,所提出的方法与利用SVM算法在传统单机平台上预测V2G功率容量进行对比分析。试验结果表明采用并行随机森林算法不仅精度比传统SVM高1.75%,速度还要快6倍多。

关键词:配电物联网;V2G功率容量;并行预测

Abstract: Accurate and fast prediction of V2G power capacity is the key to realize the energy storage of aggregated batteries of electric vehicles (EVs) as a buffer for power grid and renewable energy. However, the randomness and uncertainty of charging and discharging behavior of electric vehicles make it difficult to predict V2G power accurately. With the development of the Internet of Things (IOT) and the large-scale access of electric vehicles, the charging and discharging data of electric vehicles have increased explosively, which provides a large amount of data support for the accurate V2G power capacity, but also brings large data processing problems. A parallel prediction model of V2G power capacity is established based on the parallel random forest algorithm by using a large amount of historical and meteorological data to construct the prediction feature vector. Thus, the influence of user behavior on prediction results is avoided. In addition, a distributed big data platform based on Spark is built to realize rapid prediction of V2G power capacity. Finally, the proposed method is compared with SVM algorithm which is used in predicting V2G power capacity on traditional stand-alone platform. The experimental results show that using the parallel random forest algorithm is not only 1.75% more accurate than using the traditional SVM, but also more than 6 times faster.

Key words: distribution Internet of Things; V2G power capacity; parallel prediction

0 引言

为应对能源互联网发展对配电网的要求,基于“云大物移智”等新一代信息通信技术的配电物联网应运而生,成为能源转型“再电气化”发展的新思路、新模式、新焦点。电动汽车作为清洁环保的交通工具,近年来得到了大力推广^[1]。电动汽车所携带的电池可以作

为储能设备,为配电网提供许多辅助支撑服务^[2]。V2G功率容量预测可以为电动汽车参与电网调频、削峰填谷和有序充电等服务提供有力的数据支撑。

国内外学者V2G功率的预测已经提出多种方法,主要集中使用概率模型^[3-4]和蒙特卡洛的方法^[5]对V2G功率容量进行大致的估算。采用估算的方法,主要是由于电动汽车充放电数据的匮乏。但是,随着“云大物移智”等新一代信息通信技术的发展,可以实时获取电动汽车充放电数据,为V2G功率容量提供了新的可能。文献[6]基于Spark并行内存处理技术,利用真实的数据进行实时V2G功率容量预

收稿日期:2019-12-24;修回日期:2020-03-26

基金项目:国电南瑞科技项目(524609190013)

This work is supported by the Science and Technology Project of NARI Group Corporation(No.524609190013)

测。然而,其只是简单地认为短时间内V2G功率不变,没有充分地考虑到电动汽车接入的随机性。

目前,基于大数据的并行预测技术在电力系统中已有应用,涉及服务提升^[7]、能效管理^[8]和负荷预测等^[9-11]。其中,文献[9]基于配用电大数据进行短期负荷预测。文献[10]针对电力用户侧大规模数据,采用随机森林算法实现短期负荷并行预测。文献[11]基于灰色投影改进随机森林算法对电力系统短期负荷预测进行研究。随机森林具有很高的预测准确率,对异常值和噪声具有很好的容忍度,且不容易出现过拟合,在医学、生物信息、管理学等领域有着广泛的应用^[12],并且具有良好的并行处理能力。从而,可以利用并行随机森林算法解决V2G功率容量预测精度和速度难题。

本文基于Spark搭建了大数据平台,在考虑用户需求的基础上,确定了V2G功率容量。此外,选取不同时间节点历史数据和天气等作为预测特征变量。最终利用采集的大量历史充电数据构建日前预测训练集,基于并行随机森林算法建立日前V2G功率容量并行预测模型,并与传统SVM预测方法进行比较。

1 V2G容量并行预测方法

1.1 V2G功率容量的确定

参考文献[6],基于电动汽车当前SOC状态和用户需求,确定电动汽车是否可以参与V2G。定义V2G功率预测的时间粒度为 t_d , t_d 时间内的V2G功率容量定义如下

$$C_{t+t_d, \text{V2G}} = \sum_{i=0}^n P_{i, t+t_d, \text{V2G}} - \sum_{i=0}^m P_{i, t, \text{charge}} \quad (1)$$

式中: $C_{t+t_d, \text{V2G}}$ 为大规模电动汽车实时V2G功率容量; n 为可以参与V2G放电类的电动汽车的数量; m 为必须充电类的电动汽车的数量; $P_{i, t, \text{charge}}$ 为电动汽车 i 的充电功率; $P_{i, t, \text{discharge}}$ 为电动汽车 i 最大放电功率。

1.2 日前V2G功率并行预测

1.2.1 并行随机森林算法简介

随机森林算法(random forest, RF)定义:随机森林 f 是决策树 $\{h(X, \theta_k), k=1, 2, \dots, N_{\text{tree}}\}$ 的集合,元分类器 $h(X, \theta_k)$ 是由多个未剪枝决策树构建, θ_k 为与第 k 棵决策树独立同分布的随机向量,表示该棵树的生长过程。如图1所示。随机森林的输出是各个决策树输出结果的平均值,如式(2)所示

$$H(X^0) = \frac{1}{k} \sum_{i=1}^k h_i(X^0) \quad (2)$$

从图1可以看出随机森林算法不仅可以实现数据处理的并行化,还可以实现模型训练的并行化。

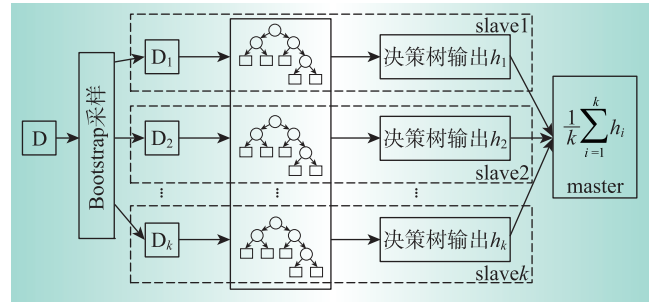


图1 随机森林算法的基本原理

Fig. 1 The basic principle of random forest algorithm

并行随机森林算法的处理过程如下:

(1) 首先利用bootstrap从分布式训练集 S 并行抽取 k 个样本,每个样本容量都与原始训练集一样。采用bootstrap取样好处是接近37%的样本不会出现在抽取的样本中,一方面可以防止出现过拟合现象,另一方面剩余的样本可以作为模型评估数据。

(2) 在各个分布式节点上,利用所抽取的 k 个样本分别并行建立 k 个决策树模型,这里为了控制模型的复杂度,设置样本最大特征数阈值 m 。

(3) 将抽样剩余的样本带入各个决策树模型中,获取各个模型的预测结果。最后,在主节点上,利用式(2)对所有的预测结果求算术平均值得到最终的回归预测结果。

(4) 判断预测结果精度是否符合要求,若不符合要求返回步骤(2)改变决策树的个数 T 和随机属性个数 m 这2个变量值重新生成模型。

1.2.2 日前V2G功率容量预测流程

日前V2G功率容量预测的时间尺度为日前24小时,同样以 t_d 为时间间隔。基于并行随机森林算法大规模电动汽车日前V2G功率容量预测流程如图2所示,主要包括3个部分:①选取特征向量构建样本训练集;②根据特征训练集利用并行随机森林算法,选取合适的决策树个数,并行建立V2G功率容量预测模型;③将预测日的特征向量带入建立好的模型得到最终的预测输出结果。

本文的预测特征向量包括3大类特征变量:天气情况、历史数据、节假日数据,如表1所示。

表1 特征向量分类展示表

Table 1 The classification display table of feature vectors

序号	类别	特征属性	变量类型	说明
1	节假日	当前时间 t	离散型	以 t_d 为间隔
2		是否是工作时间	离散型	0为否,1为是
3		是否是节假日	离散型	0为否,1为是
4		是否是周末	离散型	0为否,1为是
5	历史数据	上月平均值	连续型	对应各个时刻
6		上周平均值	连续型	对应各个时刻
7		昨日值	连续型	对应各个时刻
8	气象	温度	连续型	摄氏度
9		天气状况	离散型	0为晴天,1为非晴天

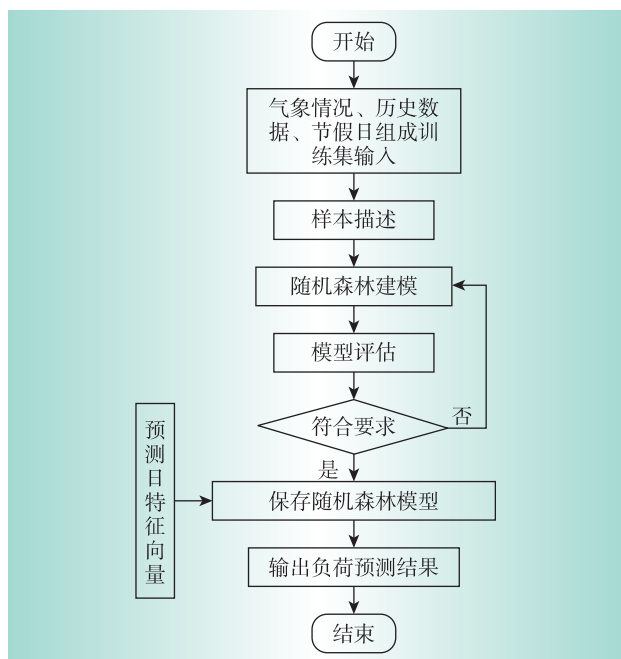


图2 大规模电动汽车功率容量预测流程图
Fig. 2 Flow chart of power capacity prediction of large-scale EVs

1.2.3 预测评价指标

采用平均相对误差 ($MAPE$) 作为大规模电动汽车功率容量预测的评价指标

$$MAPE = \frac{\sum_{i=1}^n |(\hat{y}_i - y_i) / y_i|}{n} \times 100\% \quad (3)$$

式中: $\hat{y}_i$ 为功率容量预测值; y_i 为功率容量实际值; n 为预测的点数。如果实际值为0,用V2G功率容量的平均值替换。 $MAPE$ 值越小,代表V2G功率容量预测结果越准确。

2 案例及结果分析

2.1 配电物联网架构下的样本采集

电动汽车数据采集架构如图3所示,在新型配电物联网架构下,可以通过2种方式采集电动汽车充放电数据。一种是通过智能配变终端采集充电桩的数据,然后送到云主站;另一种是电动汽车直接通过4G/GPRS上送到云主站。本文使用GPRS技术每隔1 min采集521辆电动汽车的真实充放电数据。其中包含379辆公交车,100辆出租车,37辆私家车,5辆公务车。采集了从2015年11月1日到2016年4月30日共181 d,大小为17 GB的数据。为了模拟电动汽车用户需求,在所给参数区间内随机数生成期望

的SOC,参数区间如表2所示。基于表2和式(1),最终计算出的181 d的V2G功率容量,如图4所示。

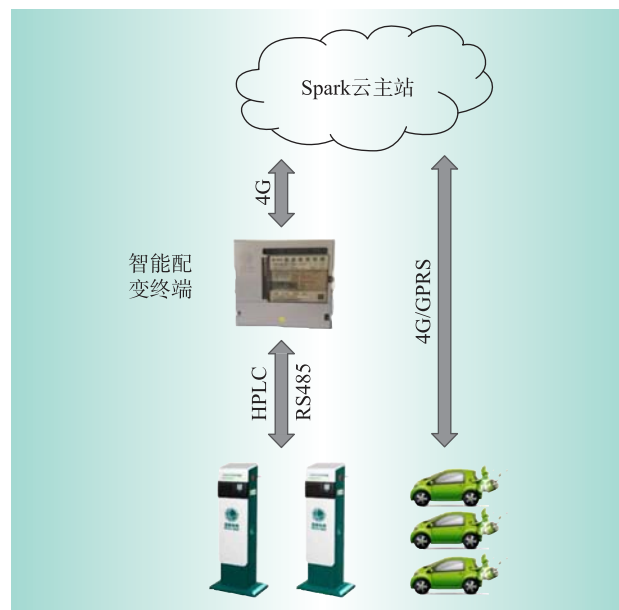


图3 电动汽车数据采集架构
Fig. 3 Data acquisition architecture for EVs

表2 用户需求参数设置

Table 2 User requirement parameter setting

类别	SOC _{min}	SOC _{max}
私家车	0.6–0.7	0.8–1
出租车	0.7–0.8	0.9–1
公交车	0.7–0.8	0.9–1
公务车	0.6–0.8	0.8–1

2.2 日前V2G功率容量预测结果分析

本文基于Spark-1.6.0搭建了1个主节点和5个从节点分布式处理平台。在服务器上虚拟6台配置相同的Ubuntu14.04虚拟机,内存为2G,CPU为2核。选用前180 d的数据作为样本训练集,最后1 d的数据作为预测集,以验证所提出的方法。采用式(3)作为预测评价指标,由于影响随机森林预测模型的精度和泛化误差的参数有决策树的个数 T 和随机属性个数 m ,通过保持一个参数不变,调整另一参数,观察对输出结果的影响。经过多次试验,最终确

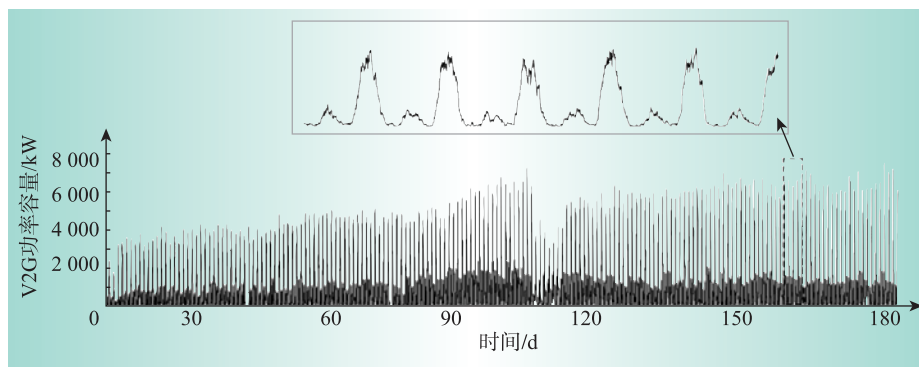


图4 181 d V2G 功率容量
Fig. 4 181-day V2G power capacity

定2个参数的值为100和3时综合效果最优,误差为7.81%,建模时间为10.1 s。从表3可以看出,采用RF预测V2G的预测精度要比传统SVM高1.75%,建模时间RF要比SVM快35.94倍,预测时间RF要比SVM快6.62倍。图5分别给出了使用RF和SVM算法的V2G功率容量预测结果和真实值对比曲线。

表3 预测模型误差对比

Table 3 Comparison of prediction model errors

算法	MPAE /%	建模时间/s	预测时间/s
RF	7.81	10.1	1.32
SVM	9.56	363	8.74

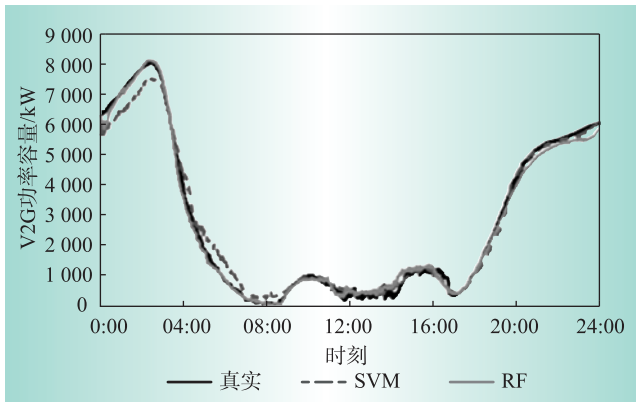


图5 预测结果对比

Fig. 5 Comparison of predicted results

3 结束语

本文基于配电物联网技术采集大量真实的电动汽车充放电数据进行日前V2G功率容量预测分析研究。利用大量历史和气象数据构建,构建了基于并行随机森林算法,建立了日前V2G功率容量的并行预测模型,并与传统SVM预测V2G功率容量进行对比分析。实验结果表明采用RF预测V2G的预测精度要比传统SVM高1.75%,建模时间RF要比SVM快35.94倍,预测时间RF要比SVM快6.62倍。D

参考文献:

- [1] 吕军,栾文鹏,刘日亮,等. 基于全面感知和软件定义的配电物联网体系架构[J]. 电网技术, 2018, 42(10): 3 108-3 115.
LV Jun, LUAN Wenpeng, LIU Riliang, et al. Architecture of distribution Internet of Things based on widespread sensing & software defined technology [J]. Power System Technology, 2018, 42(10): 3 108-3 115.
- [2] LIU H, HU Z, SONG Y, et al. Vehicle-to-grid control for supplementary frequency regulation considering charging demands. IEEE Transactions on Power Systems, 2015, 30(6): 3 110-3 119.
- [3] CHUKWU U C, MAHAJAN S M. V2G electric power ca-

- capacity estimation and ancillary service market evaluation [C]. IEEE Power and Energy Society General Meeting, 2011: 1-8.
- [4] HAN S, SEZAKI K. Estimation of achievable power capacity from plug-in electric vehicles for V2G frequency regulation: case studies for market participation [J]. IEEE Transactions on Smart Grid, 2011, 2(4): 632-641.
- [5] 李含玉,杜兆斌,陈丽丹,等. 基于出行模拟的电动汽车充电负荷预测模型及V2G评估[J]. 电力系统自动化, 2019, 43(21): 88-96.
LI Hanyu, DU Zhaobin, CHEN Lidan, et al. Trip simulation based charging load forecasting model and vehicle-to-grid evaluation of electric vehicles [J]. Automation of Electric Power Systems, 2019, 43(21): 88-96.
- [6] 岳友,陈琛,陈娜. 大规模电动汽车的实时V2G功率容量并行预测[J]. 电气自动化, 2018, 40(6): 11-13, 58.
YUE You, CHEN Chen, CHEN Na. Parallel prediction for real-time V2G power capacity of large-scale electric vehicles [J]. Electrical Automation, 2018, 40(6): 11-13, 58.
- [7] 张爽,景伟强,罗欣,等. 基于大数据的95598优质服务管理创新与实践[J]. 电力需求侧管理, 2017, 19(5): 52-54.
ZHANG Shuang, JING Weiqiang, LUO Xin, et al. Innovation and practice of 95598 quality service management based on big data [J]. Power Demand Side Management, 2017, 19(5): 52-54.
- [8] 申庆斌,武志宏,张媛,等. 基于电力大数据的用户能效服务研究[J]. 电力需求侧管理, 2017, 19(4): 29-31.
SHEN Qingbin, WU Zhihong, ZHANG Yuan, et al. Research on energy efficiency service based on big data of electric power [J]. Power Demand Side Management, 2017, 19(4): 29-31.
- [9] 丁晓,孙虹,郑海雁,等. 基于配用电大数据的短期负荷预测[J]. 电力工程技术, 2018, 37(3): 21-27.
DING Xiao, SUN Hong, ZHENG Haiyan, et al. Distribution and consumption big data based short-term load forecasting [J]. Jiangsu Electrical Engineering, 2018, 37(3): 21-27.
- [10] 王德文,孙志伟. 电力用户侧大数据分析与并行负荷预测[J]. 中国电机工程学报, 2015, (3): 527-537.
WANG Dewen, SUN Zhiwei. Big data analysis and parallel load forecasting of electric power user side [J]. Proceedings of the CSEE, 2015, (3): 527-537.
- [11] 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.
WU Xiaoyu, HE Jinghan, ZHANG Pei, et al. Power system short-term load forecasting based on improved random forest with grey relation projection [J]. Automation of Electric Power Systems, 2015, 39(12): 50-55.
- [12] 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, 2012, 26(3): 32-38.
FANG Kuangnan, WU Jianbin, ZHU Jianping, et al. A review of technologies on random forests [J]. Statistics & Information Forum, 2012, 26(3): 32-38.

作者简介:

岳友(1990),男,江苏南京人,工程师,主要从事配电网物联网技术和电能质量治理研究;

李宽宏(1992),男,福建福州人,硕士研究生,主要从事配电运维相关工作。

(责任编辑 柴明哲)