

Electricity Forecasting Based on Edge Computing under Metering System

Jun Yang¹, Aidong Xu², Wenjing Hou¹, Yixin Jiang², Hong Wen¹

¹National Key Laboratory of Science and Technology on Communications, UESTC, Chengdu Sichuan

²China Southern Power Grid Co. Ltd., Guangzhou Guangdong

Email: 1136337803@qq.com

Received: Feb. 1st, 2019; accepted: Feb. 18th, 2019; published: Feb. 25th, 2019

Abstract

Smart grid metering system has a large amount of electric power data. It can improve electric power efficiency and user experience by processing data reasonably and efficiently and making full use of data to improve the business of metering system on the user side. The edge computing model migrates all or part of the computing tasks of the application server from the cloud to the edge device on the edge of the network, thus greatly improving data transmission efficiency, and ensuring the real-time performance of data processing, while reducing the possibility of network congestion. In this paper, a method of electricity forecasting based on edge computing under metering system is proposed. This method is based on the excellent real-time performance of the edge computing model, combined with the gradient boosting decision tree algorithm and the on-line learning method in the machine learning field. It can efficiently and accurately perform real-time training and electricity forecasting on massive power data.

Keywords

Metering System, Edge Computing, Gradient Boosting Decision, Online Learning

计量系统下基于边缘计算的电量预测

杨 骏¹, 许爱东², 侯文静¹, 蒋屹新², 文 红¹

¹电子科技大学通信抗干扰国家级重点实验室, 四川 成都

²南方电网科学研究院有限责任公司, 广东 广州

Email: 1136337803@qq.com

收稿日期: 2019年2月1日; 录用日期: 2019年2月18日; 发布日期: 2019年2月25日

摘 要

智能电网计量系统拥有海量电力数据, 对数据进行合理高效的数据处理, 在接近用户侧充分利用数据进

文章引用: 杨骏, 许爱东, 侯文静, 蒋屹新, 文红. 计量系统下基于边缘计算的电量预测[J]. 智能电网, 2019, 9(1): 23-30. DOI: 10.12677/sg.2019.91003

行计量系统的业务改善,可以提高电力效率和用户体验。边缘计算模型将应用服务程序的全部或部分计算任务从云端迁移到网络边缘侧的边缘设备端执行,提高了数据传输效率,保证了数据处理的实时性,同时降低了网络拥塞的可能性。本文提出一种计量系统下的基于边缘计算的电量预测方法,该方法基于边缘计算模型优秀的实时性,结合机器学习领域的梯度提升树算法以及在线学习方式,能够高效并精准地对海量电力数据进行实时训练及进行电量预测。

关键词

计量系统, 边缘计算, 梯度提升树, 在线学习

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着云计算以及智能电网[1]系统的高速发展,海量电力数据经智能电网的计量系统[2]涌入云端进行处理。计量系统不仅为智能电网系统提供实时数据,还为云端计算提供了大量实时数据支撑。但是,大量计量终端产生的海量实时数据使云计算的性能逐渐达到了瓶颈,较长的网络延迟也使云计算模型的实时性大打折扣。同时,单纯的海量电力数据的积累并没有价值,只有通过数据挖掘、机器学习等手段,对海量电力数据进行深入分析并加以利用,才能使数据的价值发挥到最大。边缘计算[3]模型被提出并得到广泛应用。基于边缘计算模型使得数据可用在更接近计量终端进行处理,同时边缘设备的计算资源可以有效支持各种学习算法的训练,从而实现计量系统下的实时电量科学预测。利用科学有效的机器学习方法对海量电力数据进行学习,实现电量预测,为电网实时负荷提供参考数据,结合电量预测的实时电价预测可以提高电网效率、提高用户合理安全用电的用户体验[4]。

本文在计量系统下,基于边缘计算模型提出了一种基于机器学习的电量预测方法,实现了对海量电力数据的信息挖掘和对实时电量的精准预测。

2. 系统模型

2.1. 边缘计算

边缘计算模型是一种新型计算模型,旨在将应用服务程序的全部或部分计算任务迁移到网络边缘侧的边缘设备端执行,以满足实时性、隐私保护、降低能耗、快速连接等关键要求。由于原有云计算模型的大量计算任务被迁移到更靠近数据源头的附近执行,边缘计算模型极大的节省了海量数据往返云端的传输时间,提高了数据传输效率,保证了数据处理的实时性,同时降低了网络拥塞的可能性。

图1为边缘计算模型图,手机、笔记本电脑等设备产生海量数据并提交给边缘设备,边缘设备不仅执行部分计算任务,还向云端提供海量数据,同时将结果反馈到数据源头。

2.2. 基于边缘计算的电量预测模型

边缘设备相比云端更靠近数据源头,边缘计算模型下,边缘设备相对计量终端更强的计算能力,使得电量预测可以在边缘侧使用计算复杂度和精度都较高的各种学习算法。

目前的电力计量系统是通过集中器收据计量终端数据,然后上传计量管理中心,本文在计量系统下,

通过在目前的电力集中器下增加边缘计算模块，海量电力计量数据传递到集中器，边缘模块就是一个边缘设备，基于边缘模块的计算能力，以在线学习的方式结合机器学习算法，本文采用梯度提升树(GBDT)，完成对计量数据的训练和实时预测，并将实时电量预测数据返回给计量终端。其模型如图 2 所示，由于海量电力终端数据传递至边缘设备后，不再传递至云端，而是直接利用边缘设备自身的计算资源对数据进行处理，极大地提高了数据的处理效率。另外，对于每时每刻产生的新的电力数据，采用在线学习的方式可以快速地完成训练并作出预测，极大地提高了数据处理的实时性。

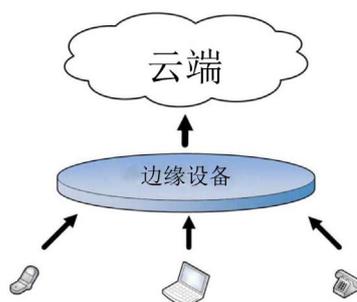


Figure 1. Edge computing model
图 1. 边缘计算模型

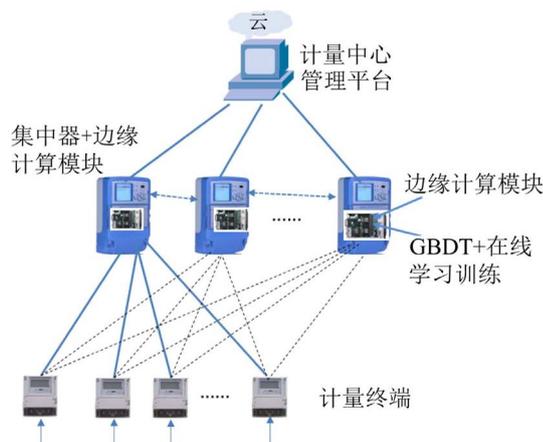


Figure 2. Electricity forecasting model based on edge computing
图 2. 基于边缘计算的电量预测模型

3. 基于梯度提升树的电量预测方法

3.1. 在线学习与离线学习

机器学习通常分在线学习与离线学习两种学习方式。

离线学习又称批量学习，是指在学习过程中，将所有或者大批量样本全部训练完毕后，才更新机器学习模型的权重。离线学习学得的模型通常是有效的，不会因某些样本的错误而使整个模型失效。但是，在实时性要求较高的任务中，离线学习方式就无法胜任。

在线学习是指在学习过程中，随着样本的输入而进行训练，单个或少量新样本的输入，会执行训练并更新模型的权重，新的模型与新的样本又会参与进下一次训练中。由于每次训练的样本数量少，所以在在线学习的速度很快，并能实时更新权重，学习到样本的变化。在线学习的方式十分适合训练无时无刻都在变化的真实世界的的数据。

3.2. 梯度提升树学习方法

梯度提升树(GBDT, Gradient Boosting Decision Tree)是一种十分有效的机器学习算法,无论是在分类还是回归问题上,都有十分优秀的表现。将 GBDT 算法应用于实时电量预测中,能够得到良好的预测效果。

在机器学习领域,梯度提升树最早由 Friedman [5]提出,属于集成学习 Boosting 家族,是其中重要的一员,并以其强悍的学习效果占据着重要地位。其以 CART 树为弱学习器,利用加法模型与前向分步算法实现学习的优化过程。

梯度提升树采用最速下降法的思想,对于损失函数:

$$L(y, f_{m-1}(x_i)) \quad (1)$$

运用最速下降法,可得下一次迭代点:

$$f_m(x_i) = f_{m-1}(x_i) - \alpha \frac{\partial L(y, f_{m-1}(x_i))}{\partial f_{m-1}(x_i)} \quad (2)$$

对于加法模型:

$$f_m(x_i) = f_{m-1}(x_i) + T(x_i; \Theta_m) \quad (3)$$

联立公式(2)(3),并取步长 $\alpha=1$ 可得:

$$T(x_i; \Theta_m) = - \frac{\partial L(y, f_{m-1}(x_i))}{\partial f_{m-1}(x_i)} \quad (4)$$

其中, $T(x_i; \Theta_m)$ 代表本次迭代需要学习的 CART 树, Θ_m 为其参数。

即利用损失函数的负梯度在当前模型的值:

$$- \left[\frac{\partial L(y, f(x_i))}{\partial f(x_i)} \right]_{f(x)=f_{m-1}(x)} \quad (5)$$

作为残差的近似值,来拟合一个 CART 树。

在计量系统下,每 30 秒采集到一个电力数据 (x_t, y_t) , $t=1,2,3,\dots$, 其中, x_t 表示第 t 次收集到的电力数据的时间等特征, y_t 表示第 t 次收集到的电力数据的电量标签,即真实电量。

以在线学习的训练方式,以当前时刻电力数据 (x_t, y_t) 为输入,下一时刻预测电力数据 (x_{t+1}, \hat{y}_{t+1}) 为输出的基于 GBDT 的电力预测方法算法步骤如下:

第一步,对于初始时刻,即 $t=0$,利用先验知识给出一个随机值 \hat{y}_1 作为第一个电量预测标签,并向计量系统返回第一个电量预测数据 (x_1, \hat{y}_1) 并输出。

第二步,对于第 t 次训练,输入训练电力数据集:

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_t, y_t)\} \quad (6)$$

并初始化模型:

$$f_0(x) = \arg \min_c \sum_{i=1}^t L(y_i, c) \quad (7)$$

其中, y_i 表示 1 到 t 个时刻的电量标签,即每个时刻的真实电量, c 表示使损失函数最小的初始化模型,即第一个 CART 树的参数。

第三步, 对 $m=1,2,\dots,M$, M 为选定的迭代次数, 分 a, b, c, d 四小步有:

a: 对每个时刻 $i=1,2,\dots,t$ 的电力数据 (x_i, y_i) , 计算当前模型负梯度:

$$r_{mi} = -\frac{\partial L(y_i, f_{m-1}(x_i))}{\partial f_{m-1}(x_i)} \quad (8)$$

b: 对 r_{mi} 拟合一个 CART 树, 得到第 m 棵树的叶节点区域 R_{mj} , $j=1,2,\dots,J$ 。这些叶节点区域 R_{mj} 会将训练数据集 T 分为 J 份。

c: 对 $j=1,2,\dots,J$, 计算:

$$c_{mj} = \arg \min_c \sum_{x_i \in R_{mj}} L(y_i, f_{m-1}(x_i) + c) \quad (9)$$

c_{mj} 即为能使当前模型的损失函数达到极小值的第 m 棵 CART 树的叶节点区域 R_{mj} 对应的输出。

d: 更新模型:

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^J c_{mj} I(x \in R_{mj}) \quad (10)$$

第四步, 得到最终模型:

$$f_M(x) = \sum_{m=1}^M \sum_{j=1}^J c_{mj} I(x \in R_{mj}) \quad (11)$$

第五步, 对下一时刻 $x_{t+1} = x_t + 30$ s, 利用最终模型计算预测电量标签:

$$\hat{y}_{t+1} = f_M(x_{t+1}) = \sum_{m=1}^M \sum_{j=1}^J c_{mj} I(x_{t+1} \in R_{mj}) \quad (12)$$

第六步, 向计量系统返回预测电量数据 (x_{t+1}, \hat{y}_{t+1}) 并输出。保存训练电力数据集 T 以及当前模型 $f_M(x)$ 。返回第二步, 进行第 $t+1$ 次训练及预测。

3.3. 基于边缘计算的电量预测流程

海量电力数据被依次传递至边缘设备, 边缘设备采用在线学习的方式, 一个电力数据或少量电力数据到来后便完成一次 GBDT 训练, 并更新一次模型, 再做出预测。

图 3 为训练流程, 由于初次预测时, 并没有任何有效数据得到训练, 所以生成一个或少量随机值作为第一个时刻或初始的某段时刻的预测值。之后, 有效的电力数据依次得到训练并获得模型, 依次对下一个时刻做出预测。

4. 实验验证

通过智能电表收集终端数据并传递至边缘计算设备, 采用本文介绍的方法获得预测结果。以每隔 30 秒为 1 个时刻, 从计量终端收集相关电力数据, 并将数据上传到边缘模块进行训练, 得到预测模型后预测下一个时刻的电量数据, 并反馈给计量系统, 并由计量系统显示出来。训练时, 以年月日时分秒为特征, 电量为标签, 在线学习训练数据量为 1 时预测效果如图 4 所示, 为 3 时如图 5 所示, 为 5 时如图 6 所示。预测电量趋势与实际获取的电量趋势相同。

图 7 为不同的预测时刻数对应的平均绝对误差[6], 由于在线学习使用的数据量不同, 所以迭代次数不同, 但预测时刻数分别为 1、3、5 对应的平均绝对误差趋势相同, 开始训练时, 数据量太少, 预测精度不高, 导致平均绝对误差迅速升高, 但经过一段时间的数据采集和训练之后, 预测模型的效果变得越来越好, 预测精度越来越高, 平均绝对误差开始下降, 且最终能趋于一个较低的平均绝对误差值。预

测时刻数为 1 时的平均绝对误差低于预测时刻数为 3 和 5，表明短期实时预测[7]的效果更好。

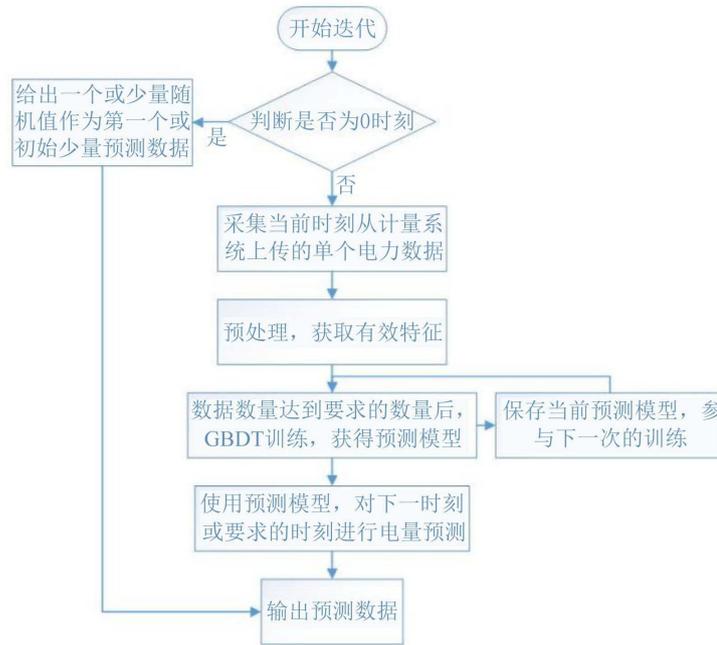


Figure 3. GBDT and online learning training process
图 3. GBDT + 在线学习训练流程



Figure 4. One-time prediction effect
图 4. 一个时刻预测效果



Figure 5. Three-time prediction effect
图 5. 三个时刻预测效果



Figure 6. Five-time prediction effect
图 6. 五个时刻预测效果

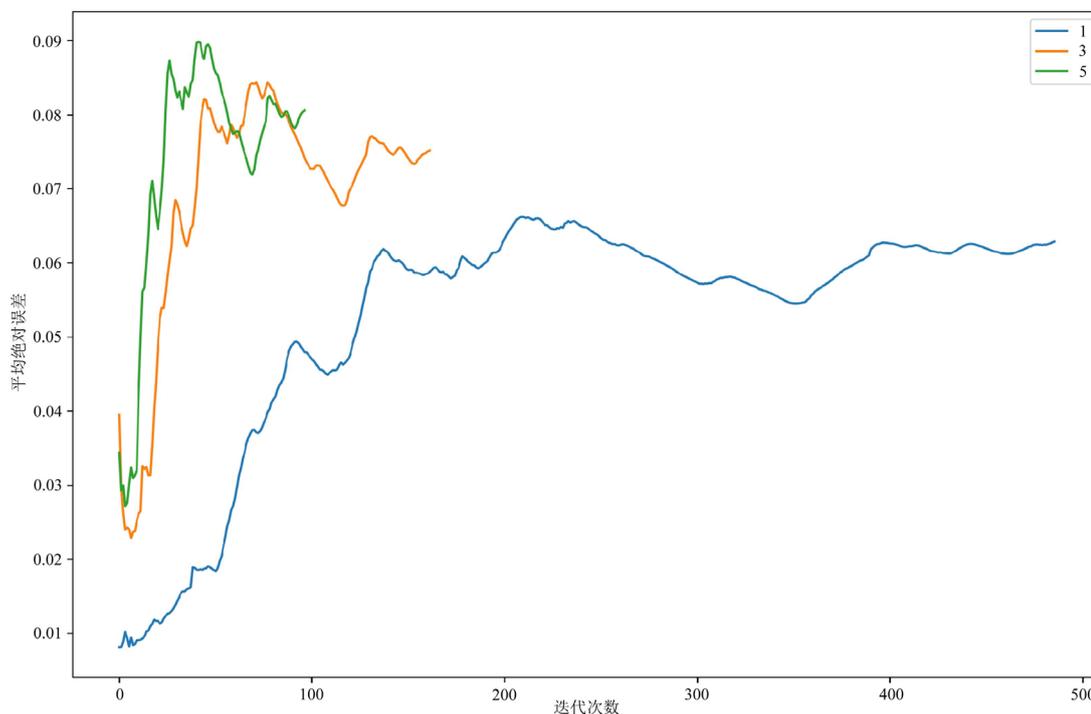


Figure 7. Average absolute error of different forecast times

图 7. 不同预测时刻数的平均绝对误差

5. 总结

随着智能电网系统的高速发展，海量电力数据涌入计量系统，合理高效的数据处理具有重大意义。本文首先介绍了边缘计算模型，机器学习算法中的 GBDT 算法以及在线学习方式，然后将其结合提出了一种在计量系统下基于边缘计算的电量预测方法，最后实验验证了方法的可行性。

基金项目

国家重大研发计划(No. 2018YFB0904900 和 2018YFB0904905)支持。

参考文献

- [1] 余贻鑫, 栾文鹏. 智能电网[J]. 电网与清洁能源, 2009, 25(1): 7-11.
- [2] 巫志平. 电能量计量系统地县一体功能需求分析与研究[J]. 智能电网, 2(5): 49-54.
- [3] 施巍松, 孙辉, 曹杰, 等. 边缘计算: 万物互联时代新型计算模型[J]. 计算机研究与发展, 2017, 54(5): 907-924.
- [4] Sousa, J.C., Neves, L.P. and Jorge, H.M. (2012) Assessing the Relevance of Load Profiling Information in Electrical Load Forecasting Based on Neural Network Models. *International Journal of Electrical Power and Energy Systems*, **40**, 85-93. <https://doi.org/10.1016/j.ijepes.2012.02.008>
- [5] Friedman, J.H. (2001) Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, **29**, 1189-1232. <https://doi.org/10.1214/aos/1013203450>
- [6] Santos, P.J., Martins, A.G. and Pires, A.J. (2007) Designing the Input Vector to ANN-Based Models for Short-Term Load Forecast in Electricity Distribution Systems. *International Journal of Electrical Power and Energy Systems*, **29**, 338-347. <https://doi.org/10.1016/j.ijepes.2006.09.002>
- [7] Duan, Q.H., Liu, J.R. and Zhao, D.F. (2017) Short Term Electric Load Forecasting Using an Automated System of Model Choice. *International Journal of Electrical Power and Energy Systems*, **91**, 92-100. <https://doi.org/10.1016/j.ijepes.2017.03.006>

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2161-8763，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：sg@hanspub.org