

基于电力交易平台大数据的多区域电价预测研究

黄靖茵¹, 黄康乾¹, 向德军¹, 刘亚², 包俊², 李呈虎²

(1. 广东电力交易中心有限责任公司, 广东广州 510600; 2. 北京清大科越股份有限公司, 北京 100084)

摘要: 随着电力市场的发展, 电力交易的实时数据通常来自于电力交易平台, 但随着电力交易量的增加, 电力交易平台用电成本逐渐增加。为此, 基于多区域电力大数据, 采用机器学习的方法, 研究了电力价格建模和预测方法, 通过分别采用线性回归、决策树和M5P三种机器学习算法, 对同一组数据集进行训练、预测和比对, 得出最佳学习算法, 最后利用预测数据与实际数据对比计算误差, 并讨论了电价预测准确性及其影响的主要因素。研究结果表明M5P算法具有最为准确的预测性能, 建立的预测模型能够准确预测最低电价以及电价变化趋势, 但是对短期电价波动的预测存在一定误差, 误差主要来源于电价自身具有的时间序列性质。

关键词: 电网大数据, 机器学习, 电力交易, 电价预测

Multi region electricity price prediction based on big data of electric power trading platform

HUANG Jingyin¹, HUANG Kangqian¹, XIANG Dejun¹, LIU Ya², BAO Jun², LI Chenghu²

(1. Guangdong Power Exchange Center Co., Ltd., Guangzhou 510600, China; 2. Qingdakeyue Corporation, Beijing 100084, China)

Abstract: With the development of power market, the real-time data of power transaction usually comes from the support of electric power trading platform, but with the increase of power transaction volume, the power cost of power data center increases gradually. Therefore, based on multi regional big data of electric power trading platform, this paper studies the modeling and forecasting methods of electric power price by using machine learning method. By using three machine learning algorithms of linear regression, decision tree and M5P respectively, the same data set is trained, predicted and compared, and the best learning algorithm is obtained. Finally, the error is calculated by comparing the predicted data with the actual data, the main factors affecting the accuracy of electricity price prediction are discussed. The research results show that M5P algorithm has the most accurate prediction performance, and the established prediction model can accurately predict the minimum price and the price change trend, but there are some errors in the short-term price fluctuation prediction.

Key words: power grid big data, machine learning, power trading, electricity price prediction

0 引言

随着电力市场的发展, 目前大部分电力交易

实时服务都建立在电力交易平台上, 以提升电力市场的可靠性和实效性。实际上电力数据中心运营成本中的30%~50%是电费支出。国际大电网会议提供的相关报告指出^[1], 仅美国的电力数据

广东电力交易中心有限责任公司科技项目资助
(GDKJXM20185842)

中心每年就要消耗910亿kW·h的电量，相当于纽约所有家庭两年的用电量。

区域实时电价是根据不同地区在一天中的时间段以及当前电力负荷来计算和制定的。因此大多数电力运营商将平台数据建立在可以提供电力市场多区域动态电价的地区，通过对比和分析电价的地域多样性，直接向电价较低的平台数据申请电力调控，从而降低运营成本。因此在电力价格可参与实时交易的背景下，电力价格预测逐渐成为了不可或缺的实用技术手段。为此研究人员在电价预测方面做了许多探索，文献[2]根据挪威电力市场的历史价格、消耗量、天气和水库数据等预测了市场电价，通过对比七种机器学习(Machine Learning, ML)算法，发现支持向量机算法预测结果最为准确，平均误差为3.14%。文献[3]提出了两种GARCH模型，基于西班牙15个月和加利福尼亚12个月的电力数据分别预测西班牙和加利福尼亚电力市场的每小时电价，结果表明采用GARCH模型的预测误差约为9%。文献[4]采用了基于聚类分析和决策树的数据挖掘技术来分析和预测巴西四个子电力市场（北方、东北、南方和中东部）的电价，结果表明对四个子市场的短期电价整体预测准确性分别为97.01%、87.87%、89.55%和94.02%。文献[5]使用已知的电价均值和方差，通过高斯随机变量对未来24 h内的电力价格进行了建模和预测。

可见针对电力价格预测，虽然目前已经取得了一定进展，但是各个国家电力市场适用的预测模型和方法都不尽相同，导致仍然还没有通用的电力价格预测方法和模型。这一方面是由于不同国家在地理位置、资源分布、电力生产、消费和政策上具有较大差异，另一方面是因为电力价格自身具有时间序列的性质^[5]，因此难以提出具有普适性的电力价格预测模型。为此本文以我国电力市场为例，基于公开的电力大数据，使用机器学习的方法建立电价预测模型，对电价波动、变化趋势进行建模预测。研究结果可为基于大数据技术的电力价格预测和建模提供参考。

1 数学模型

1.1 电价预测算法

通常利用回归类型的模型进行电力价格预测，本文采用机器学习的方式构建电价回归模型，以提高预测准确度、降低计算复杂度以及提供连续性电价预测结果，采用三种不同的机器学习算法同时对同一组数据集进行训练分析，以获得最佳算法，采用的算法如下。

(1) M5P算法

基于M5P算法^[6]构建的预测模型中，M5P算法是一种基于Quinlan M5算法的重构算法，这种算法可将常规决策树与树中节点处的线性回归函数结合在一起，从而构建预测模型。

(2) 决策树算法

决策树模型^[7]由具有层次性的结构型规则表组成，是一种典型的分类方法，首先对数据进行处理，利用归纳算法生成可读的规则和决策树，然后使用决策对新数据进行分析。

(3) 线性回归算法^[8]

根据输入的观测电价数据集，通过线性回归模型得到观测数据中自变量和因变量之间的关系。

1.2 预测效果评估模型

电价预测效果最直观的评估方法就是比较预测值与实际值之间的差异，为此采用平均绝对误差(MAE)用于评估价格预测模型的预测性能，平均绝对误差的计算公式为

$$MAE = \sum_{i=1}^n \frac{|Y_i - A_i|}{n} \quad (1)$$

式中， Y_i 为实际电价； A_i 为预测电价； n 为数据数量。

电价预测模型的构成、预测流程如图1所示，整个流程共有数据预处理、模型选择和电价预测三个主要部分。

(1) 数据预处理

数据预处理步骤的主要目的是有效降低数据维度，删除不相关的数据，提高机器学习的准确性，并提高预测结果的可靠性。

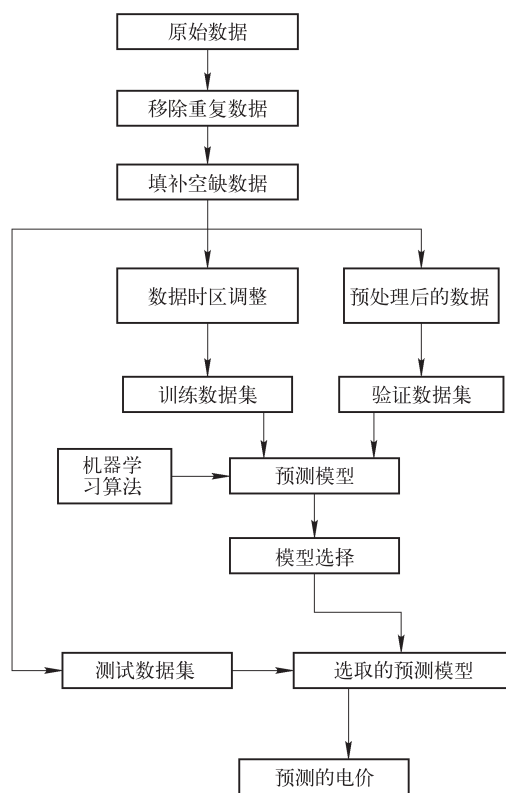


图1 电价预测流程示意图

Fig. 1 Flow chart of electricity price prediction

理想情况下用于计算的电价数据，应当包含时间、日期和每小时价格，但是实际上电价数据存在部分信息缺失、信息重复记录等现象，因此需要对数据进行预处理，比如删除重复数据，根据前后数据的逻辑关系填补缺失信息等，对统计数据进行处理，再输入计算模型。大多数电网运营商会获得动态电价的地区建立平台数据，在此过程中需要调整市场的时区，使时区统一指向北京时间。

(2) 模型选择

数据预处理之后，将数据代入到不同的机器学习（ML）算法中，通过对比选出最精确的预测模型。此后还要进行模型评估，在选定模型后，估计模型在新数据上使用产生的误差。其中选择模型和评估模型准确性的方法是将数据集分为如下三个相互独立的部分。

1) 训练数据集：用于采用机器学习算法进行训练，以建立预测模型的数据集。

2) 验证数据集：用于将预测数据与实际数

据进行比较，计算其平均误差的数据集。

3) 评估数据集：用于代入所选择的模型，进行模型准确性验证和评估的数据集。

本文基于机器的学习方法，评估和选择最佳的机器学习算法，在算法的评估和选择过程中需要应用训练数据集和验证数据集。对相同的一组训练数据和验证数据，分别代入到不同机器学习算法中，就可以生成不同的预测模型，此后就可以对不同的预测模型的预测效果进行对比和评估。在生成预测模型的过程中，共使用M5P、决策树模型和线性回归三种不同的算法，采用相同的一组训练、验证数据进行计算。最后通过比较使用这三种算法，基于同样的9组实际电价数据集构架的预测模型，计算三种模型各自的MAE值并选择最准确的模型。

(3) 电价预测

在为每个区域电力市场选择出最佳预测模型之后，将其用于电价评估，预测该区域的未来电价。最后采用针对所有市场的同期新测试数据来评估针对每个区域市场的预测模型，从而实现区域电力价格比较。

2 仿真实验结果与讨论

本文使用的实际电价数据主要来自于广东电网、福建电网以及冀北电网公开的电力数据^[9]。对每个电力市场在2018~2020年的年度数据中每小时电价（kW·h）进行了仿真实验，以建立、评估和选择最精确的电价预测模型。因此，实验共采用了9个电价数据集，即3个电力市场以及每个电力市场中各自3个年度内的电力数据集。在训练、建立和评估预测模型时，需要考虑这3个电力市场所有电力价格的特征。

对于本文所考虑的每个电力市场，利用其数据集，使用三种ML算法（M5P、决策树模型和线性回归）对市场3个年度的数据集进行测试，并且比较模型分析结果。数据集中的90%用于训练和构建预测模型，10%用于验证。模型评估和选择之后，选取2019年和2020年前两个月的数据集用于评估所选用的预测模型。

基于广东电网的电力数据,采用三种算法训练得到的预测模型的比较结果如图2所示。由图2中可以清楚地看出,由M5P算法生成的模型,其MAE数值比决策树模型和线性回归模型都要小,这一规律在2018~2020年份的数据中都有体现,而决策树模型、线性回归模型的误差则在不同年份中有不同的大小区别。

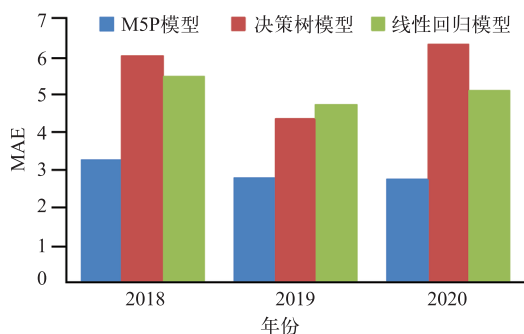


图2 广东电网电价预测误差示意图

Fig. 2 Guangdong electricity market price forecast error diagram

根据福建电网的电力数据集,采用三种算法训练得到的预测模型的比较结果如图3所示。由图3中可以清楚地看出,三种模型的预测误差从大到小分别是线性回归算法、决策树模型算法和M5P算法,表明采用M5P算法训练和建立的模型预测准确度最高。

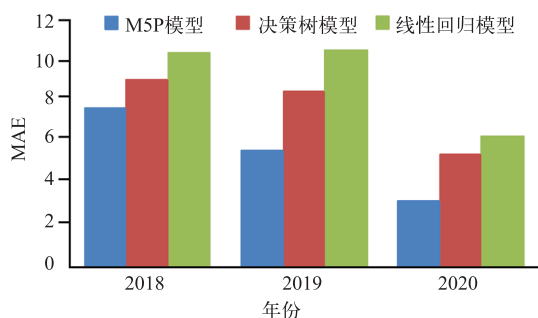


图3 福建电价预测误差示意图

Fig. 3 Fujian electricity market price forecast error diagram

基于冀北电网的电力数据,采用三种算法训练得到的预测模型的比较结果如图4所示。由图4可以清晰地看出,采用M5P算法训练和建立的模型预测准确度最高,但三种算法之间的差异并不是很大。

综合比较三种算法的应用效果,可知M5P算法效果最好,因此选用M5P算法作为机器学习算

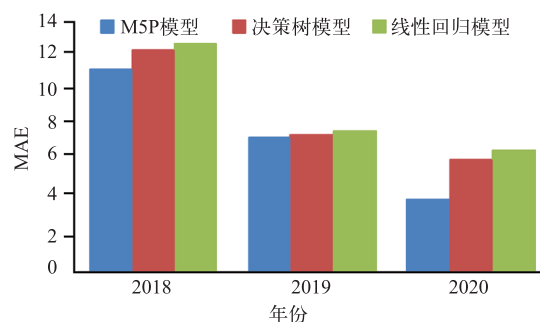


图4 冀北电网电价预测误差示意图

Fig. 4 Jibei electricity market price forecast error diagram

法进行模型构建,选取2018年的数据作为训练数据集来建立电价预测模型。通过模型预测出2019年的电价数据,再将预测数据与2019年实际的电价数据作对比,来分析预测的准确性,结果如表1所示。

表1 2019年度电力价格预测误差统计

Tab. 1 Statistical table of power price forecast error in 2019

月数	广东	福建	冀北
1月	2.88	4.75	5.75
2月	3.67	3.55	4.41
3月	6.26	4.59	6.12
4月	7.12	5.12	4.01
5月	9.16	5.97	5.48
6月	12.14	8.30	5.19
7月	5.55	11.72	5.56
8月	13.82	7.90	4.47
9月	11.21	5.22	5.63
10月	10.20	5.16	5.25
11月	7.76	4.41	6.74
12月	3.77	3.09	10.92
平均	7.79	5.82	5.79

由表1中可见,广东市场的MAE平均值为7.79,其中1月份数值最低,为2.88;8月份数值最高,为13.82。福建的MAE的平均值为5.82,其中12月出现最小值,为3.09;7月出现最大值,为11.72。冀北的MAE平均值为5.79,其中最小值为4月的4.01;最大值为12月的10.92。其中MAE出现最大值是由于电力价格飙升和价格市场波动而出现价格峰值,尽管预测模型可以反映出电力价格的正常变化趋势,但无法很好地预测价格的突然上涨和剧烈波动。针对3个电力市场的2020年前两个月的电价预测,结果如表2所示。

表2 2020年度电力价格预测误差统计

Tab. 2 Statistical table of power price forecast error in 2020

月数	广东	福建	冀北
1月	6.01	20.04	72.51
2月	10.98	6.82	19.66

2019、2020年两年中1~2月由冀北采集的电价波动示意图(共1 416 h)如图5和图6所示。可见电价从0.2元/(kW·h)飙升至0.8元/(kW·h),其中最高价格可达1元/(kW·h),对应表1冀北地区电价在2019年1月和2月的预测结果的MAE分别为5.75和4.41。观察图6可以看出,2020年前两个月的实际电力价格具有较高波动性,波动范围在0~1.05元/(kW·h)之间,对应如表2所示。根据预测结果,冀北地区在2020年1月和2月的测试MAE结果为72.51和19.66,其中2020年1月,由于价格飙升和市场波动,出现了价格的峰值,此时冀北市场上的MAE值非常高。可见预测值出现较大误差时,通常伴随着实际电力价格的大幅度波动,因此预测模型难以反映出电力价格在短期内的剧烈波动现象。

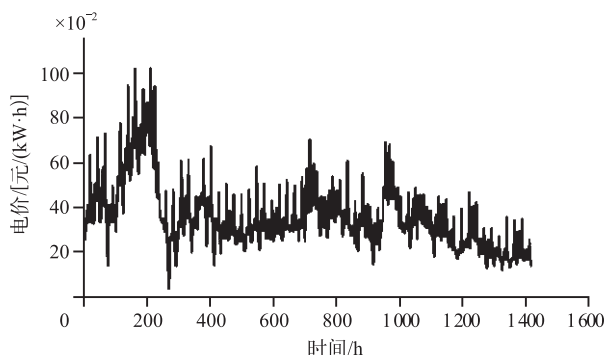


图5 冀北电网2019年电价波动示意图

Fig. 5 North Hebei electricity price fluctuation in 2019

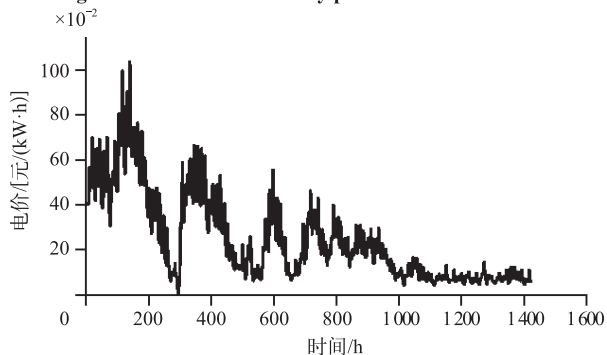


图6 冀北电网2020年电价波动示意图

Fig. 6 North Hebei electricity price fluctuation in 2020

3个电力市场2019年1月2日一天内的预测价格和实际价格的对比如图7~图9所示。由图7~图9中可以看出,虽然预测得到的价格波动与实际价格波动存在一定差异,但是预测价格很好地反映出了电价在一天内的最小值。因此实际电力交易中,运行商可以根据预测结果在不同的时间段参与不同电力调度平台的电力交易,以尽量降低运营成本。

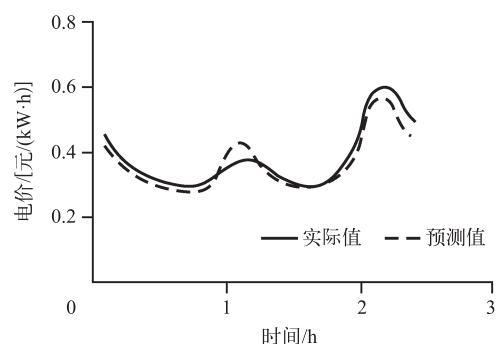


图7 广东地区一天内电价波动示意图

Fig. 7 One day price fluctuation in Guangdong area

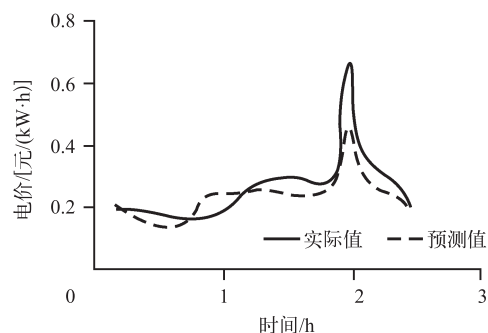


图8 福建地区一天内电价波动示意图

Fig. 8 One day price fluctuation in Fujian

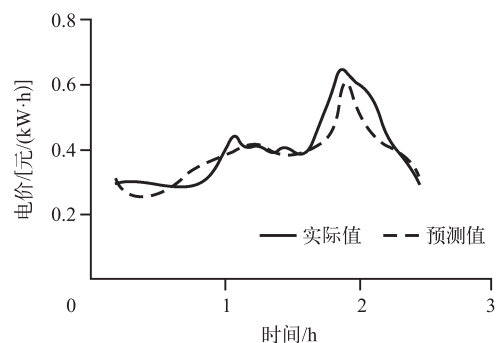


图9 冀北地区一天内电价波动示意图

Fig. 9 One day price fluctuation in North Hebei area

由于电力价格的统计特征通常会随着时间变化,例如电价均值、方差随时间并非恒定,具有

较大随机性,且通常与季节、节日等有关,要准确预测电价波动十分困难,而本文提出的预测模型成功反映出了电价的总体变化趋势。因此可认为预测模型具有较好的预测效果,考虑到电价自身的复杂时间序列特性,在电价波动上出现的预测误差是可以接受的。

3 结束语

本文基于公开的电力大数据,使用机器学习的方法建立电价预测模型,对电价波动、变化趋势进行建模预测。得到的主要结论如下:

1) 本文预测得到的电力价格很好地反映出了电价在一天内的最小值,但预测得到的价格波动与实际价格波动存在一定差异。

2) 在M5P、决策树和线性回归三种算法中,M5P算法具有最好的建模和数据训练效果,可用于建立电价预测模型。

3) 电价自身具有复杂的时间序列特性,因此要准确预测短时间内电价波动十分困难,而本文提出的模型预测出了电价整体的长期趋势,因此可认为预测模型具有较好的预测效果。

参考文献

- [1] Arens Marlene, Worrell Ernst, Eichhammer Wolfgang. Drivers and barriers to the diffusion of energy-efficient technologies—a plant-level analysis of the German steel industry[J]. Energy Efficiency, 2017, 10(2): 441-457.
- [2] SHINE P, MURPHY M D, UPTON J, et al. Machine-learning algorithms for predicting on-farm direct water and electricity consumption on pasture based dairy farms[J]. Computers & Electronics in Agriculture, 2018, 150: 74-87.
- [3] GARCIA R C, CONTRERAS J, AKKEREN M V, et al. A GARCH forecasting model to predict day-ahead electricity prices[J]. IEEE Transactions on Power Systems, 2005, 20(2): 867-874.

- [4] LAGERHOLM M, PETERSON C, BRACCINI G, et al. Clustering ECG complexes using Hermite functions and self-organizing maps[J]. IEEE Transactions Biomedical Engineering, 2000, 47(7): 838 - 848.
- [5] J C Reston Filho, C M Affonso, R C L de Oliveira. Energy price classification in North Brazilian market using decision tree[C]. 2015 12th International Conference on the European Energy Market (EEM). IEEE, 2015.
- [6] RAO L, LIU X, XIE L, et al. Minimizing electricity cost: optimization of distributed internet data centers in a multi-electricity-market environment [C]. IEEE International Conference on Computer Communications, IEEE, 2010.
- [7] Kazuya Haraguchi, Seok-Hee Hong, Hiroshi Nagamochi. Bipartite graph representation of multiple decision table classifiers[C]// Stochastic Algorithms: Foundations and Applications, 5th International Symposium, SAGA 2009, Sapporo, Japan, October 26-28, 2009. Proceedings. Springer-Verlag, 2009.
- [8] 王德文. 基于云计算的电力数据中心基础架构及其关键技术[J]. 电力系统自动化, 2012, 36(11): 15-19.
- [9] 刘碧丹. 基于SDN的电力数据中心架构研究及性能分析[D]. 北京: 华北电力大学, 2015.

作者简介



黄靖茵

黄靖茵(1992—),女,本科,助理工程师,主要研究方向为电力交易和信息技术。

(收稿日期:2019-09-26) **EA**