

基于大数据挖掘电量预测方法的创新及应用

徐 俊, 徐文辉, 曾 鑫, 宋 乐

(国网浙江省湖州供电公司, 浙江 湖州 313000)

摘 要:本文主要是基于营销系统数据开展电量预测出发,分析不同区域、行业以及时期的电量变化趋势,深入挖掘数据,发现隐藏信息并加以利用。全方位挖掘影响电量的因素,筛选出相关性强的因子形成因子储备库为后续建模做准备。通过运用时间序列、多元线性回归和灰色预测等算法对电量进行预测,丰富电量预测手段,提高短期、中期和长期的电量预测能力,并对预测结果进行可视化展示,为电量预测提供可靠的数据支撑。经实例验证,该方法能够有效的提高电量预测方面的精准度,实现未来用电量的精准预测,不仅能提升客户服务能力,提高客户服务部门的工作效率,还能有效提高电力公司的核心竞争力,使电力公司在市场愈发激烈的现状下做出更加明智的商业决策,抢占电力市场的主导地位。

关键词:营销系统;电量预测;可视化展示;数据支撑

文章编号:2096-4633(2018)10-0057-05 **中图分类号:**TM74 **文献标志码:**B

随着电力行业的业务数据不断增长,已形成了海量的客户用电数据,标志着电力企业已经进入了大数据时代。在电力大数据时代下,大数据已成为电力企业进行决策的主要依据之一。但是,单纯数据的积累并不能给电力企业带来益处,只有运用相关的数据挖掘技术手段,对大量的数据进行深入挖掘,发现隐含的信息并加以利用,进而指导电力企业做出正确的决策,这样电力大数据的作用才能发挥到极致。因此,通过基于大数据挖掘的电量预测技术,分析各供电区域、各不同行业以及各个时期的电量变化趋势,建立时间序列、多元线性回归和灰色预测等预测模型,提高短期、中期和长期电量预测能力。

1 算法理论

1.1 皮尔森算法

皮尔森相关系数是按积差方法计算,即在协方差的基础上除了两个标准差之积得出的,以两变量与各自平均值的离差为基础,通过两个离差相乘来反映两变量之间相关程度。一般用 r 来代表相关

$$X_t = \Phi_1 X_{t-1} + \cdots + \Phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

其中: $X_t = \Phi_1 X_{t-1} + \cdots + \Phi_p X_{t-p}$ 为自回归部分, $\Phi_1, \cdots, \Phi_p$ 为自回归系数, $\varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$ 为滑动平均部分, $\theta_1, \cdots, \theta_q$ 为滑动平均系数^[4]。

1.3 多元线性回归模型

在回归分析中,如果有两个及以上的自变量,则

系数,通常根据关联度 r 的绝对值大小将相关关系分为四类^[1-3],具体算法如下:

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$$

- (1) 当 $|r| < 0.3$ 时视为极弱相关;
- (2) 当 $0.3 \leq |r| < 0.5$ 时视为弱相关;
- (3) 当 $0.5 \leq |r| < 0.8$ 时视为显著相关;
- (4) 当 $|r| \geq 0.8$ 时视为强相关。

1.2 ARIMA 模型

时间序列 ARIMA 模型称为差分自回归移动平均模型,其中:AR 表示自回归,I 表示差分;MA 表示移动平均。ARIMA 模型包含三个参数: p 、 d 、 q , p 表示自回归的阶数, d 表示差分的阶数, q 表示移动平均的阶数,该模型的实质是先对非平稳的时间序列进行 d 次差分处理,形成新的平稳的数据序列,然后将平稳序列数据拟合 ARMA(p , q) 模型,最后再将原 d 次差分还原,便可以得到原序列的预测数据。其中 ARMA(p , q) 的一般表达式为:

为多元回归^[5]。多元回归模型较时间序列模型而言更容易加入较多的外部变量。我们以经济、天气等外部因素为自变量,电量为因变量,随机误差用 ε 表示,则多元线性回归方程式为:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_n x_n + \varepsilon$$

其中: x_i 为自变量, y 为因变量, b_0 为常数, b_i 为回归系数, ε 为随机误差^[6]。

1.4 灰色预测模型

对灰色系统建立的预测模型称为灰色模型,简称 GM 模型,是微分方程模型,可用于对描述对象作长期、连续、动态的反映,它揭示了系统内部事物连续发展变化的过程。灰色模型就是通过少量的、不完全的信息,建立灰色微分预测模型,对事物发展规律作出模糊性的长期描述。灰色预测需要样本量较小,运算方便,建模精度高^[7-8]。

2 实证研究

2.1 数据获取

通过对指标的研究梳理发现,台区或行业的电量数据可以从平台数据库中获取;经济类指标可以从统计局网站上获取;气象类指标可以从气象局网站上获取。

表 1 数据来源
Tab. 1 Data sources

数据类型	指标名称	数据来源
电量数据	年粒度用电量	平台数据库
	月粒度用电量	平台数据库
	日粒度用电量	平台数据库
	气象类指标	气象局网站
其他数据	节假日天数	人工统计
	春节标志	人工统计
	经济类指标	统计局网站

2.2 相关分析

从平台数据库获取电量数据进行深度挖掘,找出和电量相关性较大的因素,分别是经济类因素(国内生产总值、规模以上工业总产值、固定资产投资等)、持续假期、报装容量和气象类因素(雷电、风力和气温)。选择和开发分别适合短期、中期和长期电量预测的数据挖掘算法时加入不同的影响因素,增加电量预测准确率。



图 1 相关因素图
Fig. 1 Chart of related factor

2.2.1 经济类因素与用电量的相关性分析

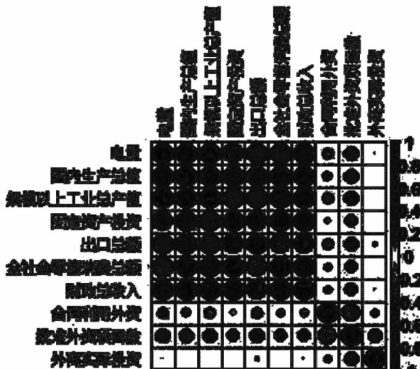


图 2 经济类因素与用电量的相关性
Fig. 2 Economic factors and electricity consumption

将经济类因素例如国内生产总值、规模以上工业总产值、固定资产投资等与用电量进行相关性分析,发现用电量与合同利用外资、批准外资项目数呈负相关,与其余经济因素呈正相关。即随着合同利用外资、批准外资项目数的增加,用电量逐渐下降;随着国内生产总值、规模以上工业总产值、固定资产投资、出口总额等经济因素的增加,用电量逐渐上升。九个经济因素中与用电量相关性较强的是国内生产总值、规模以上工业总产值、固定资产投资、出口总额、全社会零售消费总额和财政总收入,可见相比对外经济因素,用电量与国内经济因素相关性更大。

2.2.2 气温与用电量的相关性分析

表 2 气温与用电量的相关性
Tab. 2 Temperature and electricity consumption

Usage	Pearson 相关性	Usage	max	min	ave
Usage	显著性 (双侧)	1	.423 ^{**}	.458 ^{**}	.451 ^{**}
	平方与叉积的和	2.294E+17	53757916359	53911555449	53834735904
	协方差	2.896E+14	67876157.02	68070145.77	67973151.39
	N	793	793	793	793
max	Pearson 相关性	.423 ^{**}	1	.903 ^{**}	.977 ^{**}
	显著性 (双侧)	.000		.000	.000
	平方与叉积的和	53757916359	70525.306	58943.130	64734.218
	协方差	67876157.02	89.047	74.423	81.735
	N	793	793	793	793
min	Pearson 相关性	.458 ^{**}	.903 ^{**}	1	.974 ^{**}
	显著性 (双侧)	.000	.000		.000
	平方与叉积的和	53911555449	58943.130	60371.904	59657.517
	协方差	68070145.77	74.423	76.227	75.325
	N	793	793	793	793
ave	Pearson 相关性	.451 ^{**}	.977 ^{**}	.974 ^{**}	1
	显著性 (双侧)	.000	.000	.000	
	平方与叉积的和	53834735904	64734.218	59657.517	62195.868
	协方差	67973151.39	81.735	75.325	78.530
	N	793	793	793	793

** 在 .01 水平 (双侧) 上显著相关。

零假设:用电量与日最高(低、平均)温是零相关。

备择假设:用电量与日最高(低、平均)温不是零相关。

由表 2 可知,用电量与日最高温间的简单相关

系数为 0.423,与日最低温间的简单相关系数为 0.458,与日平均温间的简单相关系数为 0.451。它们的相关系数检验的概率 p 值都近似为 0。因此,当显著性水平 α 为 0.01 时,都应拒绝相关系数检验的零假设,认为用电量与日最高(低、平均)温存在线性关系,用电量受日最高(低、平均)温的正向影响。

2.2.3 持续假期与用电量的相关性分析

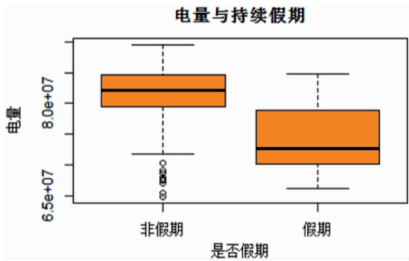


图 3 持续假期与用电量的相关性

Fig. 3 Continued holidays and electricity consumption

从用电量与持续假期的分组箱线图中可以看出,电量与持续假期相关性较高,假期与非假期相比用电量偏低。

2.3 构建模型

- 长期(1 年)电量预测→灰色预测模型
- 中期(1 月)电量预测→多元线性逐步回归模型
- 短期(7 日)电量预测→ARIMA 自动定阶模型

2.3.1 长期预测模型

部分企业的发展战略与外部经济因素息息相关,经济因素的变化对长期电量趋势产生影响,因此在开发灰色模型对长期电量进行预测时,需考虑下一阶段经济形势和宏观调控政策的情况,支撑企业年度电量调整决策^[9-10]。

灰色预测模型建立过程如下:

- (1)将用电量数据加工生成数;
- (2)对残差(模型计算值与实际值之差)修订后,建立差分微分方程模型;
- (3)基于关联度收敛进行下一步分析;
- (4)对灰色模型所得数据进行逆生成还原;
- (5)采用“五步建模(系统定性分析、因素分析、初步量化、动态量化、优化)”法,建立 GM(灰色)预测模型^[11]。

图 4 中红色三角代表真实值(即电量历史数据),蓝色曲线为拟合-预测曲线,蓝色曲线上的点代表模型拟合值,曲线最后一个点即为预测值。

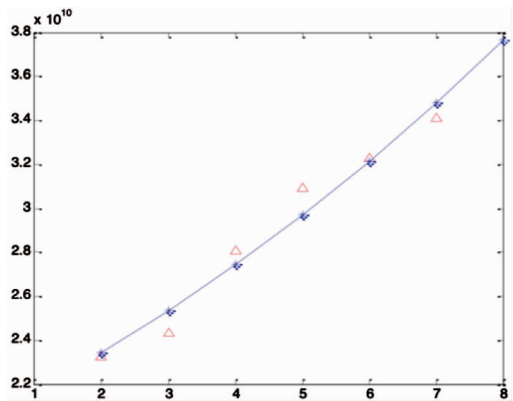


图 4 灰色模型预测示意图

Fig. 4 Schematic diagram of grey model prediction

2.3.2 中期预测模型

通过观察中期电量变化趋势发现中期电量与持续假期相关性较高,假期与非假期相比用电量偏低。持续假期包括春节、五一、十一等,在持续假期期间,大多数工厂和学校放假停工,造成用电量下降,在春节期间电量变化尤为明显。根据社会用电总计时序走势发现夏季用电量普遍高于冬季,故在选择中期电量预测模型时,采用多元线性回归模型并加入持续假期与气象类影响因素,指导电网生产,调整电网运营方式^[12-13]。

多元线性回归模型建立过程如图 5 所示。

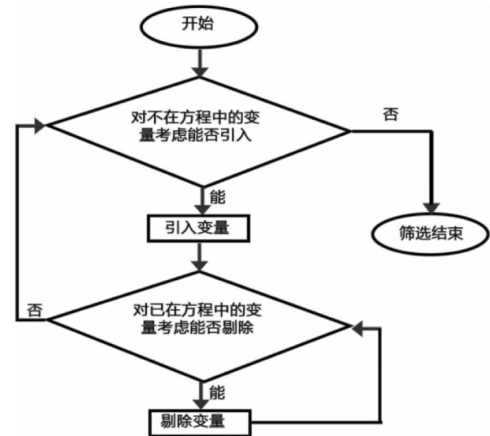


图 5 多元线性回归模型建立流程图

Fig. 5 Flow chart of multivariate linear regression model

2.3.3 短期预测模型^[14-15]

将温度分类低温、正常、高温三类,发现温度对用电量的影响显著,温度升高,用电量会随之升高(见图 6)。除了温度,业扩报装容量对短期电量亦有一定的影响。电力公司在业扩报装业务施工前制定停电计划,施工时会造成电量部分损失,短期内影响电量变化。最终在开发短期电量预测模型时,加

入温度与报装容量这两个影响因素,经过反复验证对比,选择拟合预测效果均优的 ARIMA 自动定阶模型,达到电量预测考核指标的目标值。

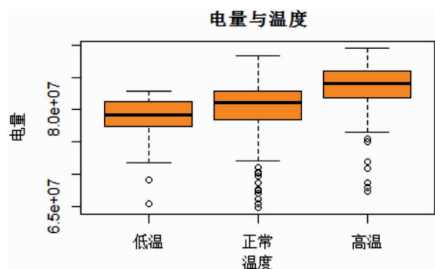


图 6 温度与电量的关系

Fig. 6 Temperature and electricity

ARIMA 模型建立过程如图 7 所示:

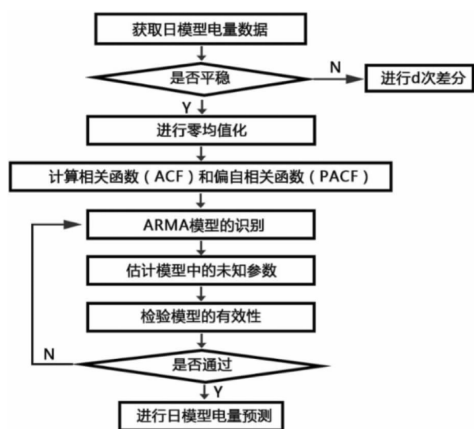


图 7 ARIMA 模型建立流程图

Fig. 7 Flow chart of ARIMA model

2.4 模型验证

以某区域为例,模型预测出该区 2016 年 7 月工业用电量为 34941.49,2016 年 8 月工业用电量为 32445.32,2016 年 9 月工业用电量为 28817.10,而实际用电量分别为 34316.55、32044.84 和 30108.93,预测误差分别为 1.82 %、1.25 % 和 4.29 %,该模型累计平均误差为 2.71 %,认为模型预测效果良好。

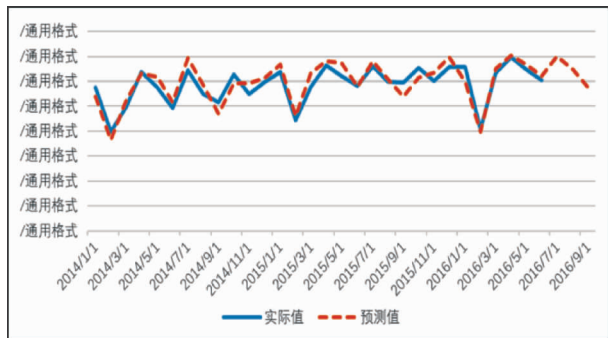


图 8 月模型拟合 - 预测图

Fig. 8 Month model fitting - prediction chart

3 应用价值

电力系统用电量预测是指从电力负荷自身的变化情况以及经济、气象等因素的影响规律出发,通过对历史数据的分析,探索事物之间的内在联系和发展变化规律,以未来的经济、气象等因素的发展趋势为依据,对电力需求作出预先的估计和推测,对于提高用电量预测的精度对增强电力系统运行的安全性和改善其经济性具有重要意义,而且在电改形势下,预测结果越准确就越能使电力公司等市场参与者在竞争且多变的环境下做出更加明智的商业决策。

4 结语

本文利用大数据挖掘的电量预测技术,结合各供电区域、各不同行业以及各个时期的电量变化趋势,并引入经济类因素数据,气候类因素数据,运用时间序列 ARIMA 模型,多元线性回归以及灰色预测模型分别建立短期电量预测,中期电量预测及长期电量预测模型,实现未来用电量的精准预测,提高对用户用电需求的管理程度,将会提高客服部门的工作效率,为实现主动、精准的客户服务提供决策支持,以提升客户服务能力。

参考文献:

- [1] HOSEINIAN FATEMEH SADAT, ABDOLLAHZADEH ALIAKBAR, REZAI BAHRAM. Semi-autogenous mill power prediction by a hybrid neural genetic algorithm[J]. Journal of Central South University, 2018, 25(01): 151 - 158.
- [2] FATEMEH SADAT HOSEINIAN, ALIAKBAR ABDOLLAHZADE, SAEED SOLTANI MOHAMADI, et al. Recovery prediction of copper oxide ore column leaching by hybrid neural genetic algorithm[J]. Transactions of Nonferrous Metals Society of China, 2017, 27(03): 686 - 693.
- [3] FEIBO JIANG, QIANWEI DAI, DONG LI. Nonlinear inversion of electrical resistivity imaging using pruning Bayesian neural networks[J]. Applied Geophysics, 2016, 13(02): 267 - 278.
- [4] 崔和瑞, 王娣. 基于季节 ARIMA 模型的华北电网售电量预测研究[J]. 华东电力, 2009, 37(01): 70 - 72.
- [5] 仇新宇, 张凌浩, 田诺. 用电量高速增长与新增用电容量的回归分析[J]. 华东电力, 2004, 32(06): 14 - 16.
- QIU Xinyu, ZHANG Linghao, TIAN Nuo. Regression analysis of rapidly increasing power consumption and newly added consuming capacity [J]. East China Electric Power, 2004, 32(06): 14 - 16.

- [6] 王文圣,丁晶,赵玉龙,等. 基于偏最小二乘回归的年用电量预测研究[J]. 中国电机工程学报,2003,23(10):17-21
WANG Wensheng, DING Jing, ZHAO Yulong, et al. Study on the load term prediction of annual electricity consumption using partial least square regressive model[J]. Proceedings of the CSEE,2003,23(10):17-21.
- [7] 董军,潘云鹤. 用神经网络预测负荷的路由选择方法[J]. 电子学报,2001,29(02):257-259
DONG Jun, PAN Yunhe. A routing approach based on load prediction using neural network[J]. Acta Electronica Sinica,2001,29(02):257-259.
- [8] 牛东晓,陈志业,邢棉,等. 具有二重趋势性的季节型电力负荷预测组合优化灰色神经网络模型[J]. 中国电机工程学报,2002,22(01):29-32
NIU Dongxiao, CHEN Zhiye, XING Mian, et al. Combined optimum gray neural network model of the seasonal power load forecasting with the double trends[J]. Proceedings of the CSEE,2002,22(01):29-32.
- [9] 欧阳森,冯天瑞,李翔,等. 基于 Odds-Matrix 算法的中长期电量组合预测方法及其应用[J]. 华南理工大学学报(自然科学版),2014,42(08):39-44
OUYANG sen, FENG Tianrui, LI Xiang, et al. Odds-matrix algorithm-based combination forecasting method of medium and long term electricity consumption and Its application[J]. Journal of South China University of Technology (Natural Science Edition),2014,42(08):39-44.
- [10] 李翔,欧阳森,冯天瑞,等. 一种基于用电行业分类的中长期电量预测方法[J]. 现代电力,2015,32(06):86-91.
LI Xiang, OUYANG Sen, FENG Tianrui, et al. A medium and long-term electricity demand forecasting method based on industry classification[J]. Modern Electric Power,2015,32(06):86-91.
- [11] 朱伟华,常虹. 基于 GRA 的优选组合技术在中长期负荷预测中的应用[J]. 煤炭技术,2013,32(01):213-215
ZHU Weihua, CHANG Hong. GRA based optimal combined technique for mid/Long term load forecasting[J]. Coal Technology,2013,32(01):213-215.
- [12] 郑雅楠,单葆国,顾宇桂,等. 中长期电量灰色预测中数据预处理方法研究[J]. 中国电力,2013,46(10):111-114+118.
ZHENG Yanan, SHAN Baoguo, GU Yugu, et al. Data preprocessing for grey model of medium-long term load forecasting[J]. Electric Power,2013,46(10):111-114+118.
- [13] 邵丹,林辉,郝志峰,等. 考虑春节影响的中期电量预测[J]. 电机与控制学报,2007,11(05):555-558.
SHAO Dan, LIN Hui, HAO Zhifeng, et al. Forecasting method of medium-term electricity consumption under the impact from the Spring Festival[J]. Electric Machines And Control,2007,11(05):555-558.
- [14] 张冰玉. 基于数据挖掘技术的短期电力负荷预测[J]. 电力大数据,2017,20(10):18-21.
ZHANG Bingyu. Prediction of short-term power load based on data mining technology[J]. Power Systems and Big Data,2017,20(10):18-21.
- [15] 张然然,刘鑫屏. 火电机组超短期负荷预测[J]. 热力发电. 2018,48(07):52-57
ZHANG Ranran, LIU Xinping. Ultra-short-term load forecasting for thermal power units[J]. Thermal Power Generation,2018,48(07):52-57.

收稿日期:2018-08-20

作者简介:



徐 俊(1977),男,本科,工程师,主要从事电力营销管理工作。

(本文责任编辑:范 斌)

Innovation and application of power forecasting method based on big data mining

XU Jun, XU Wenhui, ZENG Xin, SONG Le

(1. State Grid Zhejiang Huzhou Electric Power Supply Company, Huzhou 313000 Zhejiang, China)

Abstract: This paper is based on the marketing system data to conduct electricity forecasting, analyze the trends of electricity consumption in different regions, industries and periods, drill down into the data, find hidden information and use it. The factors affecting electricity consumption are excavated in all directions, and the relevant factor formation factor reserve library is selected to prepare for subsequent modeling. Through the use of time series, multiple linear regression and gray prediction algorithms to predict electricity, enrich power forecasting methods, improve short-term, medium-term and long-term power forecasting capabilities, and visualize the forecast results to provide reliable data support for power forecasting. The example proves that the method can effectively improve the accuracy of power forecasting and achieve accurate prediction of power consumption in the future, which not only improves customer service capability, improves the work efficiency of customer service departments, but also effectively improves the core competition of power companies, so that power companies can make more wise business decisions in the increasingly fierce market situation and seize the leading position in the electricity market.

Key words: marketing system; electricity prediction; visual display; data support