

· 负荷预测专题 ·

基于配用电大数据的短期负荷预测

丁晓¹, 孙虹², 郑海雁², 季聪², 徐金玲³, 仲春林², 熊政²

(1. 国网江苏省电力有限公司, 江苏南京 210024; 2. 江苏方天电力技术有限公司, 江苏南京 211102; 3. 国网江苏省电力有限公司电力科学研究院, 江苏南京 211103)

摘要: 受限于数据信息的不完整和粗粒度, 短期网供负荷预测的准确率一直难以进一步提升, 而配用电信息系统数据的积累和大数据技术的快速发展为开展基于配用电大数据的短期负荷预测提供了数据基础和技术支撑。本文首先针对配用电大数据的特征分析了“脏数据”的来源与类型, 并提出了相应的数据清洗方法; 其次基于大量的历史负荷、电量和气象数据, 构建了行业负荷温度影响模型和行业电量节假日影响模型; 最后基于上述用电影响模型开展了江苏电网短期网供负荷预测, 实际计算结果验证了预测效果的有效性和准确性。

关键词: 配用电大数据; 数据清洗; 负荷温度影响模型; 电量节假日影响模型; 短期负荷预测

中图分类号: TM732

文献标志码: A

文章编号: 2096-3203(2018)03-0021-07

0 引言

随着配电网信息化建设的推进, 配电网在日常运行中产生了大量的配用电数据, 但一直以来这些数据并未得到充分的挖掘和有效的利用。如今电改政策试点、售电侧放开对电力客户服务提出了更高的要求, 电力行业市场化进程的深入也对电力负荷预测提出了更高的要求。目前, 国内外专家和学者已经在大数据负荷预测领域展开了研究工作, 也取得了一些成果。文献[1-3]对配用电大数据背景下用电预测场景进行了研究和分析, 认为用电预测是配用电大数据的关键技术和重要应用之一。文献[4]提出了一种基于并行随机森林的大数据负荷预测方法, 取得了比决策树更高的预测精度。文献[5]将局部加权线性回归预测算法和云计算 Map/Reduce 模型相结合, 用于开展短期电力负荷预测研究。文献[6]提出了基于用户的电网负荷预测方案, 将大数据技术引入到负荷预测。文献[7]采用基于大数据的神经网络算法将发电预测精度提高到接近 99%。

江苏居住区配电一体化系统的全面建成、用电信息采集系统(下称“用采系统”)的全面覆盖, 积累了自 2009 年以来全省 47 万配变、26 万专变、3700 万用户的负荷和电量数据, 营销系统保存着自 2009 年以来全省 26 万大用户的业扩报装、增容、减容数据, 江苏省电力公司气象信息系统积累了自 2006 年

以来全省 13 地市 71 个气象站的 10 min/点温度、湿度、雨量、风速等气象数据, 上述数据总量已累计达到 180 TB, 且仍然在以每日 30 GB 的速度快速增长。如何充分利用这些数据资源, 挖掘负荷、电量、业扩、气象、经济等因素的关系, 建立更加精准的负荷和电量影响模型, 提高短期负荷预测的精确度, 是本文的重点研究内容。

本文分析了大数据负荷预测方法的优势, 介绍了配用电大数据的清洗方法, 构建了多维负荷和电量模型, 实现了基于配用电大数据的短期负荷预测方法, 并且结合实际计算结果, 验证了方法的准确性。

1 大数据负荷预测方法的优势

传统负荷预测方法大致可以分为统计算法和智能算法, 统计算法包括时间序列模型、决策树、回归算法、随机森林等, 智能算法包括人工神经网络、支持向量机、贝叶斯理论等基本算法及其改进算法, 目前组合预测算法也受到了业界的青睐和关注^[8-9]。但上述方法由于建模时选取的样本较小, 历史数据的选取直接影响负荷预测的效果。大数据负荷预测方法存在以下 3 点优势。

(1) 考虑的影响因素更全。影响负荷走势的因素众多, 主要包括两大类型: 用户用电行为中体现的随机性, 以及外部气象因素和节假日的影响^[10-11]。大数据负荷预测方法掌握着全面的用电行为数据和外部气象数据, 可以通过聚类分析挖掘用户用电行为模式, 对不同类型的用户用电行为特征进行详细的分析, 可以通过相关性分析挖掘负荷走势受气象因素影响程度。

收稿日期: 2017-12-19; 修回日期: 2018-01-28

基金项目: 国家重点研发计划资助项目(2016YFB0901100); 国家电网有限公司科技项目“提升电力营销服务能力的大数据关键技术研究”

(2) 数据的时间跨度更长。大数据负荷预测方法选取了时间跨度更长的历史数据,用于发现负荷数据随月、季、年周期发生的变化规律,与选取相似日^[12]负荷数据进行预测的一些传统方法相比,不存在相似日选取不当而造成预测误差较大的问题。

(3) 数据的空间粒度更细。大数据负荷预测方法所采用的负荷数据粒度可以细化到地区、行业、变压器、线路、台区、用户等各个级别,而由于江苏地域广阔,存在温度差异(南北最高温差可达 20℃),行业负荷特征明显,因此本文的大数据负荷预测方法采用分地区、行业预测再按比例叠加的思路。这种方法能够考虑不同行业的负荷特性、不同地区的温度差异,而且也便于对预测结果进行成分分析和误差追踪。

2 数据源建设

江苏电力大数据平台以营配集成、用电信息采集、省地县一体化电量系统为基础,结合外部气象和经济数据,建成了江苏配用电大数据中心,为江苏配用电大数据分析工作提供了丰富的数据资源。

2.1 配用电大数据的来源与分类

配用电大数据中心的数据体系架构如图 1 所示。

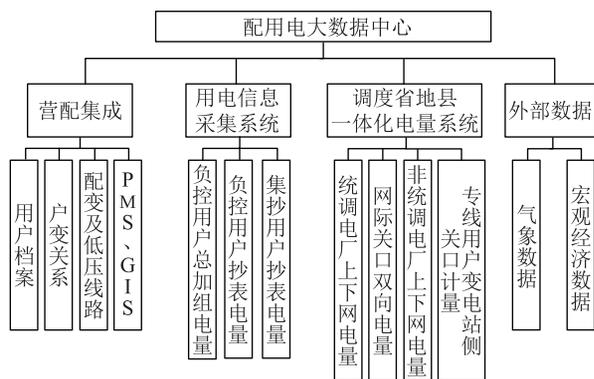


图 1 配用电大数据中心的数据体系架构

Fig.1 Structure of distribution and consumption big datacenter

数据主要来源于营配集成、用电信息采集系统、调度省地县一体化电量系统,以及外部的气象数据和宏观经济数据。

(1) 营配集成数据。全省 3596 万居民用户档案信息、47 万配变及其与用户的户变关系、配网相关低压线路档案信息、电力生产管理系统(power production management system, PMS)和地理信息系统(geographic information system, GIS)数据。

(2) 用采系统数据。全省 3596 万居民用户、26 万负控用户自 2009 年以来的负荷和电量数据。

(3) 调度省地县一体化电量系统数据。自 2005 年以来全省约 2500 个厂站的超过 40 000 只电表的上下网电量数据和关口计量数据。

(4) 外部数据。自 2006 年以来全省 13 地市 71 个气象站 10 min/点的温度、湿度、雨量、风速等气象数据,以及江苏 13 地市 54 县的宏观经济数据。

2.2 配用电大数据的预处理

从大数据平台目前集成的所有数据类型来看,“脏数据”主要有 3 种大类型,11 个小类,如表 1 所示。

表 1 “脏数据”类型
Tab.1 Type of dirty data

“脏数据”类型	详细分类
数据缺失	负荷数据缺点
	电量数据为空
	温度数据为空
数据异常	负荷数据异常(跳大数、跳小数)
	负荷数据为负值
	负荷数据超容
	电量数据异常(跳大数、跳小数)
	电量数据为负值
数据不同步	温度数据异常
	负荷数据采集频率不同(24 点、48 点、96 点)
	负荷数据与温度数据频率不同(温度 10 min/点,负荷 15 min/点)

数据异常从某种程度上来说,与数据缺失的结果是类似的,因此针对需预处理的 3 种数据类型,分别提出了两种不同的数据清洗方法。

2.2.1 数据缺失/异常的清洗方法

数据缺失/异常的清洗主要采用了替代法和插值法。

(1) 插值法。负荷数据缺点(异常)较少时,可以基于当日负荷曲线,采用插值法(如拉格朗日插值、三次样条插值等)实现负荷曲线的补全;电量数据缺点(异常)较少时,可以基于当月电量曲线,采用插值法实现电量数据补全。

(2) 替代法。负荷数据缺点(异常)较多,无法采用插值法时,可以用相似日(工作日选取上一周工作日,周末选取上一周周末)同一时段负荷数据替代;电量数据缺点(异常)较多时,可以用相似日(工作日选取上一月工作日,周末选取上一月周末)的电量数据替代;温度数据为空时,处理方法类同负荷数据缺点。

2.2.2 数据不同步的清洗方法

数据不同步的情况下,通常采用平均值法、强制同步法进行数据清洗。

(1) 平均值法。由于极少部分终端采集频率为 48 点/日,因此需要将 48 点负荷数据扩展为 96 点负荷数据,可以采用平均值法(如 8:45 的数据取为 8:30 和 9:00 的平均值)进行数据扩展。

(2) 强制同步法。温度数据为 10 min/点,而负荷数据 15 min/点,强制将 00:10 的温度数据与 00:15 的负荷数据匹配,00:30 的温度数据与 00:30 的负荷数据匹配,00:40 的温度数据与 00:45 的负荷数据集匹配,以此类推。

3 多维用电影响因素模型的构建

3.1 模型构建的总体思路

用电量受气象因素、节假日、经济形势等众多因素的影响,因此筛选合适的影响因素加以分析,建立精确的用电影响因素模型,是分析用电特性、开展用电预测的基础。由于经济数据发布频率太低,而且经济环境在一段时间内相对于气象因素而言比较稳定,因此本文只考虑气象因素和节假日建立用电影响因素模型。用电影响因素模型的总体构建思路如图 2 所示。

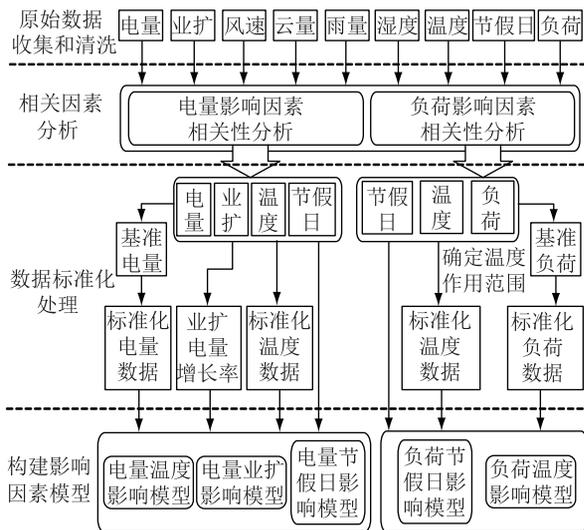


图 2 用电影响因素模型构建思路

Fig.2 Construction of electricity consumption impact model

模型构建思路主要包括以下 4 个步骤:

(1) 原始数据收集和清洗。收集电量、负荷数据,历史气象信息和节假日数据,按 2.1~2.2 所述方法处理电量和负荷数据,按 2.3 处理气象数据,根据历史节假日前后电量数据适当增加节假日的长度(如春节假期规定 7 d,但实际上春节前一周、后两周很多行业都处于停工状态),得到清洗后的各项数据。

(2) 相关因素分析。逐一计算气象因素、节假

日与电量、负荷的相关系数,得到与电量、负荷相关性最高的影响因素。

(3) 数据标准化处理。获取(1)中清洗好的温度、电量(负荷)数据,计算得到各温度档位下的平均电量(负荷),绘制温度电量(负荷)影响曲线,将电量(负荷)走势较为平缓的温度区间作为基准电量(负荷)温度区间。

(4) 用电影响模型构建。以 $^{\circ}\text{C}$ 为单位,将各温度区间下的电量(负荷)进行归类并计算平均值,对于负荷模型还要考虑分时特性,得到电量温度影响模型、负荷温度影响模型等。

3.2 相关因素分析

用电量受气象因素、节假日、经济形势等众多因素的影响,由于经济数据发布频率太低,而且经济环境在一段时间内相对于气象因素而言比较稳定,因此这里只考虑气象因素和节假日与用电量的相关性。

目前气象信息考虑温度、湿度、雨量、云量、气压、风速六项指标,采用式(1)的相关性计算方法分别对各影响因素进行分析^[13-16]:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

式中: x_i 为用户第*i*天的日用电量数据; \bar{x} 为用户*n*天的日用电量平均值; y_i 为第*i*天的影响因素数据(例如温度、湿度等); \bar{y} 为*n*天的影响因素数据平均值。

一般认为, $|r| \geq 0.8$ 视为两个向量高度相关, $0.3 \leq |r| < 0.8$ 时中度相关, $|r| < 0.3$ 时不相关。以南京为例,选取 2014 年南京全年用电量、气象和用电量数据,进行相关性分析。需要说明的是:

(1) 由于温度较高时与用电量呈正相关性,而温度较低时与用电量呈负相关性,因此选取 25°C 及以上和 10°C 以下两种情况分别进行相关性分析。

(2) 节假日信息按工作日=1,周末=2,法定假日=3 设定。

南京用电影响因素与用电量相关性计算结果如表 2 所示。

可见用电量与温度高度相关,与节假日中度相关,而与风速、湿度、雨量、云量、气压不相关,因此构建用电影响因素模型时,主要考虑温度和节假日两个关键因素。

3.3 数据标准化处理

在构建用电影响因素模型之前,需要通过计算

表2 用电量与影响因素的相关性计算结果

Tab.2 Correlation calculation results of electricity consumption and influencing factors

影响因素	<i>r</i>	影响因素	<i>r</i>
温度(≥ 25℃)	0.843 9	温度(≤ 10℃)	-0.682 8
湿度	0.031 2	气压	-0.008 6
雨量	-0.015 9	风速	-0.185 1
云量	-0.048 3	节假日	-0.320 1

选取合适的基准电量(负荷),实现电量(负荷)数据的标么化,便于后期直观地分析各影响因素对电量(负荷)的影响率。数据标准化处理主要包括以下几个步骤:

(1) 按度划分温度区间,将各温度区间对应的电量(负荷)数据归并,得到各温度区间内的平均电量(负荷)。

(2) 绘制电量(负荷)-温度曲线,并采用七点平滑算法平滑曲线。

(3) 按点计算(2)中曲线斜率,选择曲线中较为平缓的温度区间,计算该温度区间内的平均电量(负荷),作为基准电量(负荷)。

(4) 采用(3)中的基准电量(负荷),标准化所有电量(负荷)数据。

图3、4分别为南京商业和苏州居民的电量-温度曲线。

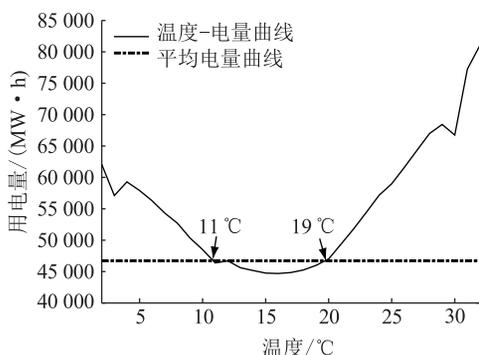


图3 南京商业电量-温度曲线

Fig.3 Electricity consumption-temperature curve of Nanjing Commercial Industry

可见南京商业在 11~19℃ 之间用电量较为稳定,温度小于 11℃ 或者大于 19℃ 时用电量逐渐上升;苏州居民在 8~23℃ 之间用电量较为平稳,在小于 9℃ 时急剧略有上升,而大于 23℃ 时用电量急剧上升,说明苏州居民用电对高温天气非常敏感。

3.4 行业用电影响模型

由 3.1 可知,用电影响模型包括电量温度影响模型、电量业扩影响模型、电量节假日影响模型、负荷温度影响模型和负荷节假日影响模型,限于篇

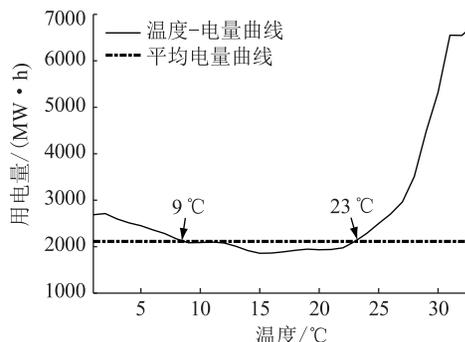


图4 苏州居民电量-温度曲线

Fig.4 Electricity consumption-temperature curve of Suzhou resident

幅,这里主要介绍行业负荷温度影响模型和行业电量节假日影响模型的构建方法。

3.4.1 行业负荷温度影响模型

(1) 首先根据 3.3 的计算方法得到待计算行业的基准负荷(基准负荷为全年工作日 96 点负荷平均值) \bar{P}_i , 其中 i 取值为 1~96。

(2) 逐日逐点计算负荷影响率:

$$R_{(d,i)} = (P_{(d,i)} - \bar{P}_i) / \bar{P}_i \quad (3)$$

式中: d 表示工作日编号, $R_{(d,i)}$ 为第 d 个工作日第 i 个点的负荷影响率。

(3) 将温度划分为 >40、<-4、-4~40 这 45 个档位,将所有工作日的 96 点负荷影响率归类到对应的温度档位,形成 45×96 的温度-负荷影响率序列 $S_{(d,i,t)}$, 其中下标 t 为温度标签。

(4) 逐一对待 $S_{(d,i,t)}$ 中的数据集合求平均值,得到温度综合影响率 $C_{(i,t)}$,若 $S_{(d,i,t)}$ 中某一格数据样本太少,则温度范围上下扩展 1℃,重新计算温度综合影响率,若果数据样本依然过少,则将该点的温度综合影响率交给后续的模式拟合算法完成。

(5) 形成负荷-温度综合影响率矩阵 $C_{(i,t)}$ 后,通过插值法修补残缺数据点,通过平滑算法平滑模型中的异常数据点,最终得到负荷温度影响模型。

(6) 由于负荷数据更新较快,且过于久远的历史数据不具备参考价值,因此负荷温度影响模型每月根据新增数据更新。

图 5 为某地区住宿餐饮业的负荷-温度影响模型。可见该行业分时负荷特性为白天负荷较高、午饭和晚饭时间出现高峰、下午略降;该行业的温度影响率为温度较低、较高时影响率较高,说明炎热/寒冷天气下用于制冷/取暖负荷较高,而温度适宜时影响率较低。

3.4.2 行业电量节假日影响模型

目前节假日包括下列七类:元旦、清明、劳动、

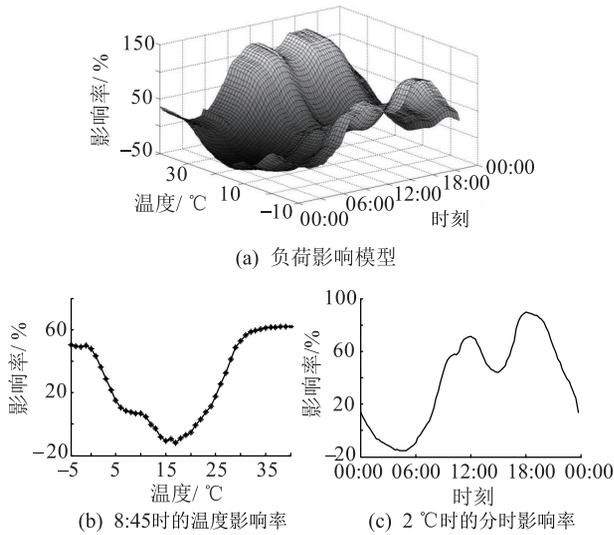


图5 某地区住宿餐饮业负荷-温度影响模型

Fig.5 Load-temperature impact model of accommodation and catering industry

端午、国庆、中秋、春节。以年为计算周期,计算每年所有节假日期间,行业日电量相对于节假日前正常电量的影响率,其计算流程如下:

(1) 根据实时节假日放假时间及调休安排,配置节假日信息表,为了显示节假日对电量的连续影响趋势,应在实际节假日的基础上前后多配置1 d,对于春节这个特殊节假日,前后多配置一周。

(2) 找节假日前最近5个工作日,计算这5个工作日的平均用电量,将该电量作为基准电量。

(3) 根据下式计算节假日期间每天的电量影响率:

$$R_d = (A_d - \bar{A}_{d-5}) / \bar{A}_{d-5} \quad (4)$$

其中: R_d 为节假日第 d 天的行业电量影响率; A_d 为节假日第 d 天的行业用电量; \bar{A}_{d-5} 为节假日前的5个工作日的行业平均用电量。

图6为某地区商业的电量节假日影响模型。

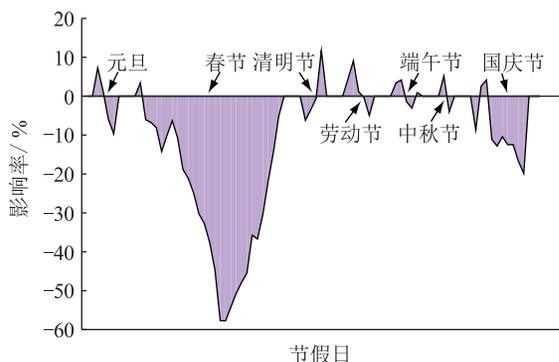


图6 某地区商业电量节假日影响模型

Fig.6 Electricity-holiday consumption impact model of commercial industry

可见春节、清明节、国庆节都对商业产生了负影响,主要是因为春节、国庆长假归乡、旅游人数较多,而清明节更多人回归家庭,也一定程度上影响了商业的用电量。

4 短期负荷预测的实现与应用

由于江苏全省用户数量高达4000万,若全省网供负荷预测分解过细(到用户)工作量太大,且用户负荷随机性较强,预测精确度反而较低。实践表明,将全省网供负荷分解到行业级即可得到令人满意的精确度,且计算量也在合理的范围内。图7为基于配用电大数据的短期网供负荷预测方法。

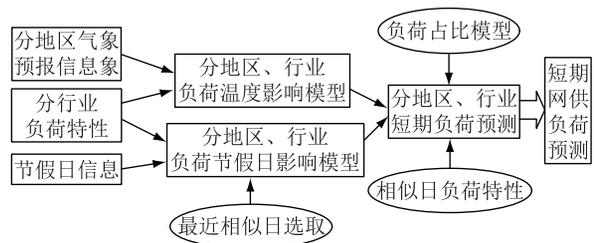


图7 基于配用电大数据的短期网供负荷预测方法示意图

Fig.7 Sketch of short-term load forecasting method based on distribution and consumption big data

(1) 最近相似日选取。周二~周五以其前一日为最近相似日,周六、周日、周一以上周同一天为最近相似日,周休日以上周周休日为最近相似日(部分行业周休日并非周六、周日)。

(2) 相似日气象因素剔除。从历史气象库中匹配得到行业所在地区最近相似日的温度数据,对照3.4.1的行业负荷温度影响模型,按下式剔除行业相似日负荷中的温度因素:

$$P_{i1} = P_i / (1 + C_{(i,t)}) \quad (5)$$

式中: i 取值范围为1~96; P_i 为相似日的第 i 个有功负荷; $C_{(i,t)}$ 为第 i 个负荷数据对应温度为 t 时的温度综合影响率; P_{i1} 为剔除气象因素后的相似日负荷。

(3) 预测日气象因素加成。从预测气象库中匹配得到行业所在地区预测日的温度数据,对照3.4.1的行业负荷温度影响模型,按下式向(2)式中的 P_{i1} 加成影响因素:

$$P_{i2} = P_{i1} (1 + C_{(i,t)}) \quad (6)$$

其中 P_{i2} 即为加成预测日气象影响因素的负荷数据。

(4) 节假日因素考虑。对于节假日,还需要基于行业电量节假日影响模型进行修正,具体方法同步骤(2)(3)中的气象因素修正方法。

(5) 构建负荷占比模型。由于统计口径的差异

和统计误差的存在,分行业、地区负荷汇总到网供负荷时会与调度口径数据存在一定的差异,因此根据相似日分行业、地区负荷和全省网供负荷,按比例得到全网负荷的占比模型。

(6) 实现网供负荷预测。根据负荷占比模型和(4)中得到的分行业、地区负荷数据,汇总得到全省网供负荷预测结果。

在传统方法中,误差逆向传播神经网络(back propagation, BP)算法应用广泛、适应性强,以BP算法为传统方法的代表,与本文提出的大数据方法进行比较。图8为BP算法和大数据方法的全省网供负荷预测误差率。

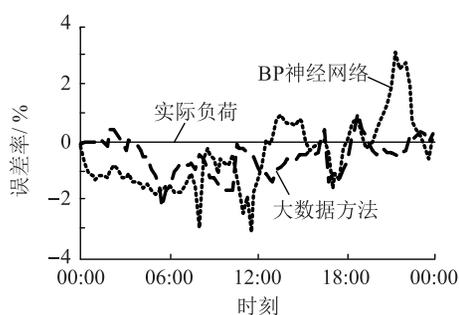


图8 短期网供负荷预测结果

Fig.8 Result of short-term load forecasting

BP算法的平均预测误差率为1.038 2%,而大数据方法的平均预测误差率仅为0.623 9%,且96个预测点中,有61个点误差率低于BP算法。可见,基于大数据的网供负荷预测方法可以与传统方法优势互补,共同提升负荷预测的准确率。

5 结语

本文基于配用电大数据开展了大量的研究工作,主要进行了:(1)配用电大数据的清洗。基于配用电大数据的特点以及实际业务的需要,分析了配用电大数据中“脏数据”的来源和类型,针对性地提出了数据清洗方法。(2)基于配用电大数据,构建了行业负荷温度影响模型和行业电量节假日影响模型,为后期开展短期负荷预测打下基础。(3)提出了基于大数据的短期负荷预测方法。基于多维用电影响因素模型,开展了分地区、行业的短期网供负荷预测,计算结果表明基于配用电大数据的网供负荷预测有着较高的准确性,可以为电网运行和规划提供数据支撑。

参考文献:

[1] 赵腾,张焰,张东霞,等.智能配电网大数据应用技术与前景分析[J].电网技术,2014,38(12):3305-3312.
ZHAO Teng, ZHANG Yan, ZHANG Dongxia, et al. Application technology of big data in smart distribution grid and its prospect

analysis[J]. Power System Technology, 2014, 38(12): 3305-3312.

[2] 刘科研,盛万兴,张东霞,等.智能配电网大数据应用需求和场景分析研究[J].中国电机工程学报,2015,35(2):287-293.
LIU Keyan, SHENG Wanxing, ZHANG Dongxia, et al. Big data application requirements and scenario analysis in smart distribution network[J]. Proceedings of the CSEE, 2015, 35(2): 287-293.

[3] 田世明,杨增辉,时志雄,等.智能配用电大数据关键技术研究[J].供用电,2015,32(8):12-18.
TIAN Shiming, YANG Zenghui, SHI Zhixiong, et al. Research on the key technology of big data for smart power distribution and utilization[J]. Distribution and Utilization, 2015, 32(8): 12-18.

[4] 王璟,杨德昌,李猛,等.配电网大数据技术分析与典型应用案例[J].电网技术,2015,39(11):3114-3121.
WANG Jing, YANG Dechang, LI Meng, et al. Analysis of big data technology in power distribution system and typical applications[J]. Power System Technology, 2015, 39(11): 3114-3121.

[5] 张素香,赵丙镇,王风雨,等.海量数据下的电力负荷短期预测[J].中国电机工程学报,2015,35(1):37-42.
ZHANG Suxiang, ZHAO Bingzhen, WANG Fengyu, et al. Short-term power load forecasting based on big data[J]. Proceedings of the CSEE, 2015, 35(1): 37-42.

[6] 王德金.基于大数据技术的短期负荷分析与预测[J].华东电力,2014,42(10):2007-2010.
WANG Dejin. Short-term load analysis and forecast based on big data technology[J]. East China Electric Power, 2014, 42(10): 2007-2010.

[7] RAHMAN M N, ESMAILPOUR A. An efficient electricity generation forecasting system using artificial neural network approach with big data [C] // IEEE First International Conference on Big Data Computing Service and Applications (Big Data Service). Redwood, CA: IEEE, 2015: 213-217.

[8] 程建东,杜积贵.组合预测方法在电力负荷预测中的应用[J].江苏电机工程,2011,30(6):38-40,44.
CHENG Jiandong, DU Jigui. Application of combined method in power load forecasting [J]. Jiangsu Electrical Engineering, 2011, 30(6): 38-40, 44.

[9] 廖旋焕,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.
LIAO Nihuan, HU Zhihong, MA Yingying, et al. Review of the short-term load forecasting methods of electric power system [J]. Power System Protection and Control, 2011, 39(1): 147-152.

[10] 张翠芝,智明.泰州电网负荷特性分析及负荷预测[J].江苏电机工程,2011,30(4):45-47.
ZHANG Cuizhi, ZHI Ming. Load characteristics and load Forecasting of Taizhou Power Grid [J]. Jiangsu Electrical Engineering, 2011, 30(4): 45-47.

[11] 许琦,曾凌.南京电网负荷与气温敏感性分析[J].江

- 苏电机工程,2012,31(5):55-57,61.
- XU Qi, ZENG Ling. Analysis on load and temperature sensitivity of Nanjing Grid[J]. Jiangsu Electrical Engineering, 2012,31(5):55-57,61.
- [12] 黎灿兵,李晓辉,赵 瑞,等. 电力短期负荷预测相似日选取算法[J]. 电力系统自动化,2008,32(9):69-73.
- LI Canbing, LI Xiaohui, ZHAO Rui, et al. A novel algorithm of selecting similar days for short-term power load forecasting[J]. Automation of Electric Power Systems, 2008, 32(9):69-73.
- [13] 贺春光. 夏季电网最大负荷与气温的相关性分析[J]. 华北电力技术, 2011, 30(5):5-17.
- HE Chunguang. Correlation analysis on system peak load and temperature[J]. Hebei Electric Power, 2011, 30(5):15-17.
- [14] 刁赢龙,盛万兴,刘科研,等. 大规模配电网负荷数据在线清洗与修复方法研究[J]. 电网技术, 2015, 39(11):3134-3140.
- DIAO Yinglong, SHENG Wanxing, LIU Keyan, et al. Research on online cleaning and repair methods of large-scale distribution network load data[J]. Power System Technology, 2015, 39(11):3134-3140.
- [15] 柯人观,周金辉,汪东辉. 微电网规划体系研究[J]. 浙江电力, 2016, 35(2):23-26.
- KE Renguan, ZHOU Jinhui, WANG donghui. Study on planning system of microgrid[J]. Zhejiang Electric Power, 2016, 35(2):23-26.
- [16] 兰 岚. 配电线路故障衍化分析[J]. 浙江电力, 2016, 35(4):28-31.
- LAN Lan. Analysis of fault evolution in distribution lines[J]. Zhejiang Electric Power, 2016, 35(4):28-31.

作者简介:



丁 晓

丁 晓(1973—),男,本科,研究员级高级工程师,从事电力系统营销管理,电力营销大数据等方面工作(E-mail: dingx@ js. sgcc. com. cn);

孙 虹(1971—),男,硕士,研究员级高级工程师,从事电力系统环保技术,电力系统信息化等方面工作;

郑海雁(1979—),男,硕士,研究员级高级工程师,从事电力系统信息化,电力营销大数据等方面工作。

Distribution and Consumption Big Data Based Short-term Load Forecasting

DING Xiao¹, SUN Hong², ZHENG Haiyan², JI Cong², XU Jinling³, ZHONG Chunlin², XIONG Zheng²

(1. State Grid Jiangsu Electric Power Co., Ltd., Nanjing 210024, China;

2. Jiangsu Frontier Electric Technology Co., Ltd., Nanjing 211102, China;

3. State Grid Jiangsu Electric Power Co., Ltd. Research Institute, Nanjing 211103, China)

Abstract: Limited to incompleteness of power load, electricity and other relevant data, the accuracy of network supply short-term load forecasting is hard to enhance. The data accumulation of distribution and utilization system and rapid development of big data technology provide data basis and technical support for big data based load forecasting. Firstly, the sources and types of 'dirty data' are analyzed based on the characteristics of distribution and utilization big data, and the corresponding methods of data cleaning are put forward in this paper. Secondly, based on a large amount of historical power load, electricity consumption and meteorological data, industry load-temperature impact model, and industry electricity-holiday impact model are established. Finally, short-term load forecasting is carried out based on above models, and test cases show effectiveness and accuracy of proposed big data based short-term load forecasting method.

Key words: distribution and utilization big data; data cleaning; load-temperature impact model; electricity-holiday impact model; short-term load forecasting

(编辑 陈 娜)