

基于自适应组合模型在供电公司电能表需求预测的分析和应用

熊紫腾¹, 李伟志²

(1. 国网江西省电力有限公司供电服务管理中心, 江西 南昌 330096; 2. 远光软件股份有限公司, 福建 福州 350000)

摘要:智能电能表用表需求采用人工上报时, 受限于主观经验, 存在偏差大、精准度低等问题。文中提出了基于自适应组合模型的电能表需求预测法, 通过构建预测模型特征参数, 开展特征重要性分析, 确定电能表需求预测影响因素的特征值, 采用基于XGBoost与随机森林模型的自适应组合预测方法, 比单一的算法模型有较好的适配性, 充分考虑了计量业务类型, 结合最佳融合权重, 可灵活匹配相对应的算法, 以得到最优的预测模型。验证结果表明, 电能表需求预测综合误差为7.98%, 有效提高了公司电能表需求响应的精准度和速度, 助力公司智能量测体系建设和精益化管理水平的提升。

关键词:智能电能表; 自适应; 需求预测; 组合模型

中图分类号: TM 933.4 **文献标志码:** B **文章编号:** 1006-348X(2024)06-0036-04

0 引言

省公司级计量中心负责全省智能电能表的采购到货、检定检测和物流配送等相关工作。当前全省智能电能表的用表需求主要通过线下收集、汇总所辖地市公司提报数量, 同时结合各单位库存情况、历年装拆数量、到期轮换数量等关联因素, 基于人工管理经验来进行确定。人工上报受限于业务人员的主观经验, 缺乏科学依据, 线下收集工作量繁重, 易造成偏差, 缺乏精准度, 容易造成电能表短缺或需求过剩等问题, 不符合计量物资精细化管理的要求。

因此, 文中开展智能电能表需求预测研究分析工作, 综合智能电能表需求相关因素, 优化需求预测模型, 从而支撑“到、检、配”计划的科学合理制定, 助力国网公司智能量测体系建设和精益化管理水平提升。

1 电能表需求预测影响因素分析

1.1 需求影响因素

计量管理工作共涉及多个业务层级和专业部

门, 电能表需求数量与其历史业务数据规律、城市规划建设、气候环境变化等内外部事物之间存在相互影响、相互制约、相互印证的关系^[1]; 同时, 不同业务对电能表需求量和响应及时性不同, 如故障抢修对用表及时性要求较高, 批量新装对需求量要求较高。因此, 需要对各类业务进行分类分析, 寻找不同业务类型之间的特征量, 为电能表需求预测提供准确的数据支撑。

综合电能表需求预测影响因素的分析, 从数据的可获取性和完整性角度出发, 主要从电能表需求预测的4个主要影响因素进行分析^[2]。

1) 电能表需求具有年、季等周期性特点。在外部环境没有大的改变时, 如政策、自然灾害等影响, 应当具有显著的周期变化特点且变化连续, 一般不会出现大的跃变。因此, 不同月份可能影响电能表需求量, 去年同期的需求量可能也会对当月需求量产生影响, 历史同期需求量平均值对预测当月需求也具有一定的参考价值。

2) 受季节、温度、天气以及节假日等影响, 不同的季节、地区会对电能表的需求量造成明显的影响。每年电能表需求量可能会发生随机波动, 但总体年

收稿日期: 2024-07-26

作者简介: 熊紫腾(1991), 男, 硕士, 工程师, 从事电能计量技术工作。

内占比与分配应当相差不大,因此将年初到本月总需求量在全年需求量的占比作为一个影响因素,年内占比最大值、最小值、平均值、中位数等也可以作为年内占比的参考因素,本月总需求量与去年需求量的比值也能反映当年需求量的水平,是否会整体增大或减小。

3) 不同类型的电能表安装量及安装趋势相差较大。分别针对不同类型的电能表,建立相应的预测模型,能提高总体预测精度。

4) 随着时间的推移,经济、气候温度等外部条件与预期有一定差异。电能表需求量及趋势可能发生变化,因此,离现在越近时间的需求量可能参考性越小。

1.2 需求预测影响因子分析

1) 构建影响因子

对电能表需求影响因素进行综合分析,挑选关键因素进行特征量分析,作为电能表需求预测的数据基础。本文主要构建安装数量、周期时长、需求平均值、同期需求量等影响因素字段,作为模型训练的输入值。具体影响因素明细如表1所示。

表1 需求预测影响因子

| 名称 | 含义 |
|--------------------|--------------|
| quantity | 当月安装量 |
| period | 历史记录距离当前时间长度 |
| during_year_ratio | 同期需求占全年比例平均值 |
| last_year_quantity | 去年同期需求量 |
| max | 同期最大值 |
| min | 同期最小值 |
| median | 所有同期平均值 |
| mean | 所有同期中位数 |
| cumsum_ratio | 当年累计量与去年同期比值 |
| month_x | x月 |

2) 相关性分析

对不同影响因子与电能表需求量建立相关性分析,找到可能影响电能表需求的因素及影响程度大小。由于电能表需求是连续型变量,为了分析变量间的趋势相关性,采用 Spearman 软件进行相关系数分析^[3]。

原始数据 x_i, y_i 按从大到小排序, x_i, y_i 为原 x_i, y_i 在排序后数据所在的位置,则 x_i, y_i 称为变量 x_i, y_i 的秩次,则 $d_i = x_i - y_i$ 为 x_i, y_i 的秩次之差。Spearman 相关系数计算公式为:

$$\rho_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (1)$$

对电能表需求预测的影响因子进行 Spearman 相关系数分析,结果如图1所示。

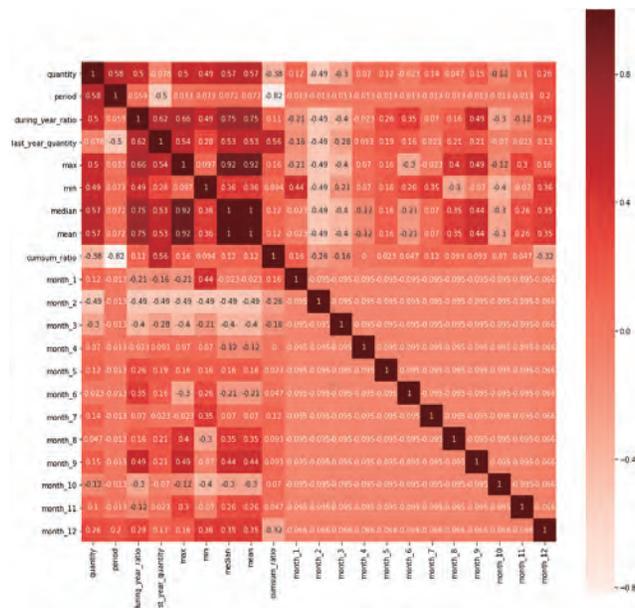


图1 影响因子相关性

通过分析图1中各个电能表需求预测的影响因子相关性,可以发现:

(1) 所有因素均与电能表需求量具有不同程度的相关性。例如,月份相关性最强的依次是2月、3月、12月,2月负相关最强代表全年在当月相对稳定偏小,12月正相关最强代表全年在当月相对稳定偏大;此外,历史数据与当前时间的差距显著影响了当月需求量。

(2) 不同影响因子之间具有不同程度的相关性。例如同期平均值与所有同期中位数相关系数为1,表示两者在给定的数据中对结果值的支撑度相同,模型入参时可考虑去除其中之一;如最大值与同期平均值、最大值与同期中位数的相关性系数均为0.92,说明三者在给定的数据中对结果值的支撑度相近,模型入参时可考虑三选一。综上相关性分析,可简化模型特征参数,降低模型训练复杂度。

2 电能表需求预测模型

2.1 XGBoost预测模型

1) XGBoost模型理论

XGBoost^[4]属于机器学习中的高级树模型,并采

用了集成学习中的 Boosting 策略, 是对提升树的改进, 它是将许多树模型集成在一起, 形成一个很强的分类器, 所用到的树模型则是 CART 回归树模型。该算法思想就是不断地添加树, 不断地进行特征分裂来生长一棵树, 每次添加一个树, 其实是学习一个新函数, 去拟合上次预测的残差。当训练完成得到 k 棵树, 要预测一个样本的分数。根据这个样本的特征, 在每棵树中会落到对应的一个叶子节点, 每个叶子节点就对应一个分数, 最后只需要将每棵树对应的分数相加就是该样本的预测值。

2) XGBoost 模型预测结果

采用 XGBoost 模型预测方法, 用基于历史数据的 XGBoost 模型对某供电公司的 2022 年单相电能表需求进行预测, 结果如表 2 所示。

表 2 XGBoost 模型预测值与实际值

| 月份 | 实际值 | 预测值 | 误差值 |
|-----|--------|--------|---------|
| 1月 | 24 111 | 30 120 | 24.92% |
| 2月 | 8 422 | 9 128 | 8.38% |
| 3月 | 15 235 | 8 990 | -40.99% |
| 4月 | 17 009 | 21 318 | 25.33% |
| 5月 | 16 801 | 24 993 | 48.76% |
| 6月 | 17 115 | 14 440 | -15.63% |
| 7月 | 26 411 | 22 983 | -12.98% |
| 8月 | 16 453 | 10 798 | -34.37% |
| 9月 | 22 333 | 29 014 | 29.92% |
| 10月 | 19 875 | 15 215 | -23.45% |
| 11月 | 26 110 | 36 223 | 38.73% |
| 12月 | 26 110 | 24 593 | -5.81% |

从表 2 可以看出, 预测最大误差为 3 月 48.76%, 最小误差为 12 月 5.81%, 综合误差 25.5%, 预测误差偏差大, 且波动较大。

2.2 随机森林预测模型

1) 随机森林模型理论

随机森林^[5]是一种集成算法, 它采用了集成学习中的 Bagging 策略, 通过组合多个弱分类器, 最终结果通过投票或取均值, 使得整体模型的结果具有较高的精确度和泛化性能。它可以取得不错成绩, 主要归功于“随机”和“森林”, 一个使它具有抗过拟合能力, 一个使它更加精准。

2) 随机森林模型预测结果

采用随机森林模型预测方法, 用基于历史数据的随机森林模型对某供电公司的 2022 年单相电能表需求进行预测, 结果如表 3 所示。

表 3 随机森林模型预测值与实际值

| 月份 | 实际值 | 预测值 | 误差值 |
|-----|--------|--------|---------|
| 1月 | 24 111 | 27 121 | 12.48% |
| 2月 | 8 422 | 7 216 | -14.32% |
| 3月 | 15 235 | 18 702 | 22.76% |
| 4月 | 17 009 | 19 809 | 16.46% |
| 5月 | 16 801 | 12 458 | -25.85% |
| 6月 | 17 115 | 19 781 | 15.58% |
| 7月 | 26 411 | 30 247 | 14.52% |
| 8月 | 16 453 | 20 494 | 24.56% |
| 9月 | 22 333 | 26 830 | 20.14% |
| 10月 | 19 875 | 24 542 | 23.48% |
| 11月 | 26 110 | 22 911 | -12.25% |
| 12月 | 26 110 | 31 676 | 21.32% |

从表 3 可以看出, 采用随机森林模型预测方法, 预测最大误差为 25.85%, 最小误差为 12.48%, 综合误差 18.34%, 误差偏差较大。

3 自适应组合预测模型分析

3.1 自适应组合预测模型

1) 特征重要性分析

针对安装数量、安装占比、安装平均值、同期需求量等影响因素关键特征量, 通过机器学习算法对特征重要性进行分析, 得到特征重要性结果, 如表 4 所示。其中, “当年累计量与去年同期比值”特征值为 0.32, 表明该特征量是预测未来电能表需求量最重要的参考因素; “历史记录距离当前时间长度”重要性次之, 特征值为 0.25, 表明越是近期的历史数据参考价值越大; 此外, “历史同期最大值”也是一个非常重要的信息。

通过特征重要性分析, 在模型入参训练时, 可按特征值降序顺序, 依次重点选择数值较大的特征量输入模型训练, 强化模型特征工程构造科学性, 减少模型训练干扰项, 提高模型预测准确性。

表 4 特征重要性分析值

| 特征量 | 特征值 |
|--------------|------|
| 当月安装量 | 0.21 |
| 历史记录距离当前时间长度 | 0.25 |
| 同期需求占全年比例平均值 | 0.19 |
| 去年同期需求量 | 0.04 |
| 同期最大值 | 0.21 |
| 同期最小值 | 0.13 |
| 所有同期平均值 | 0.01 |
| 所有同期中位数 | 0.01 |
| 当年累计量与去年同期比值 | 0.32 |
| 月份 | 0.05 |

2) 组合自适应模型优化

分析计量业务对电能表需求的影响,从业务角度出发,影响不同安装类型电能表需求的因素存在差异性。将计量业务分为改造类、业扩类和故障更换类三大类别,不同安装类型的电能表需求之间相对独立。预测模型的优化方案采用XGBoost与随机森林模型融合预测,接下来需要计算出最佳融合权重。

假设 θ_1 是模型结果值 x_1 的参数, θ_2 是模型结果值 x_2 的参数, $h_{\theta(x)}$ 为模型组合预测结果值:

$$h_{\theta(x)} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 = \sum_{i=0}^n \theta_i x_i = \theta^T x \quad (2)$$

假设 ε 为组合模型预测结果值 $h_{\theta(x)}$ 与实际值 $y^{(i)}$ 之间存在的差异,即:

$$y^{(i)} = \theta^T x^{(i)} + \varepsilon^{(i)} \quad (3)$$

其中误差的发生具有独立性,且相同分布,服从均值为0、方差为 σ^2 的高斯分布,即

$$p(\varepsilon^{(i)}) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\varepsilon^{(i)})^2}{2\sigma^2}\right) \quad (4)$$

将 $\varepsilon^{(i)} = y^{(i)} - \theta^T x^{(i)}$ 带入式(4),得到:

$$p(y^{(i)}|x^{(i)}; \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right) \quad (5)$$

利用对数似然方程解决输入 θ_1, θ_2 为多少时误差 $\varepsilon^{(i)}$ 为最小,即梯度下降求最优解:

$$\begin{aligned} \log L(\theta) &= \log \prod_{i=1}^m \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right) = \\ &= \sum_{i=1}^m \log \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right) = m \log \frac{1}{\sqrt{2\pi}\sigma} - \\ &= \frac{1}{\sigma^2} \cdot \frac{1}{2} \sum_{i=1}^m (y^{(i)} - \theta^T x^{(i)})^2 \end{aligned} \quad (6)$$

目标让上述对数似然函数越大越好,即对如下目标函数求偏导,得到最小值:

$$J(\theta) = \frac{1}{2} \sum_{i=1}^m (y^{(i)} - \theta^T x^{(i)})^2 = \frac{1}{2} (X\theta - y)^T (X\theta - y) \quad (7)$$

当偏导等于0时,

$$\theta = (X^T X)^{-1} X^T y \quad (8)$$

最终,利用计算得出的各模型权重参数值 θ ,加权平均计算得出最终预测值。

3) 自适应组合预测模型验证

结合上述特征分析及模型权重进行优化调整,采用自适应组合预测模型算法,对某供电公司的2022年单相电能表需求进行预测,结果如表5所示。

表5 自适应组合预测模型预测值与实际值

| 月份 | 实际值 | 预测值 | 误差值 |
|-----|--------|--------|---------|
| 1月 | 24 111 | 26 050 | 8.04% |
| 2月 | 8 422 | 9 439 | 12.08% |
| 3月 | 15 235 | 14 117 | -7.34% |
| 4月 | 17 009 | 18 669 | 9.76% |
| 5月 | 16 801 | 18 011 | 7.20% |
| 6月 | 17 115 | 18 219 | 6.45% |
| 7月 | 26 411 | 28 030 | 6.13% |
| 8月 | 16 453 | 15 709 | -4.52% |
| 9月 | 22 333 | 23 900 | 7.02% |
| 10月 | 19 875 | 21 822 | 9.80% |
| 11月 | 26 110 | 22 822 | -12.59% |
| 12月 | 26 110 | 27 750 | 6.28% |

从表5可以看出,采用组合模型预测方法,预测最大误差为12.59%,最小误差为4.52%,综合误差7.98%,误差范围满足业务需求。

4 结语

1) 开展特征重要性分析,构建电能表需求预测的特征影响因素,通过机器学习算法,得到各特征参数的特征值,其中“当年累计量与去年同期比值”特征值为0.32,表明该特征量是预测未来电能表需求量最重要的参考因素,其次“历史记录距离当前时间长度”和“历史同期最大值”。

2) 相较单一的算法模型,自适应组合模型有较好的适配性,充分考虑了计量业务类型,可灵活匹配相对应的算法,以得到最优的预测模型,并结合网格搜索来确定最佳融合权重,最终输出组合模型的预测结果。通过自适应组合预测模型验证结果表明,电能表需求预测综合误差为7.98%。

3) 构建基于业务特征的组合电能表需求预测模型,实现计量器具近期及中远期需求的精细化管控,改变了传统用表需求以人工提报的工作方式,推进对计量器具的均衡分布和高效调度,提高了供电公司电表需求响应的精准度和速度,有效强化公司安全生产保障能力和电能计量公共服务能力。

参考文献:

- [1] 彭楚宁,杜新纲,李天阳,等.基于业务特征的智能电能表需求预测模型研究[J].电测与仪表,2019,56(3):139-143.
- [2] 李兵,李翀,吴一敌,等.Holt-Winters模型在电能表需求预测中的应用[J].河北电力技术,2019,38(5):7-9.(下转第61页)

2) 汽轮机猫爪翘起的本质是汽缸稳定性被破坏,汽缸稳定性主要受到自身重力、管道接口推力、汽流对汽缸反作用力等因素的综合影响。

3) 通过减小管道冷拉力及优化调整支吊架受力,降低了管道对超高压缸的推力,超高压缸猫爪翘起现象得到了彻底根治,汽轮机振动情况良好,汽轮机效率得到了明显提升。该研究成果为类似机组猫爪翘起、汽缸膨胀不畅等疑难问题的解决提供了较好的参考作用。

参考文献:

- [1] 国家能源局.火力发电厂管道支吊架验收规程:DL/T 1113—2024[S].北京:中国电力出版社,2024:9-10.
- [2] 中华人民共和国国家质量监督检验检疫总局,中国国家标准化管理委员会.管道支吊架:第1部分技术规范:GB/T

- 17116.1—2018[S].北京:中国标准出版社,2018:19-20.
- [3] 陈盛广,王军民,邓玲惠,等.高温再热蒸汽管道异常位移产生原因及防治研究与应用[J].热力发电,2022,51(5):169-174.
- [4] 国家能源局.发电厂汽水管道应力计算技术规程:DL/T 5366—2014[S].北京:中国电力出版社,2014:23-28.
- [5] 国家能源局.火力发电厂汽水管道设计规范:DL/T 5054—2016[S].北京:中国电力出版社,2016:33-40.
- [6] 中华人民共和国住房和城乡建设部.电厂动力管道设计规范:GB/T 50764—2012[S].北京:中国计划出版社,2012:104-114.
- [7] 国家质量技术监督局,中华人民共和国建设部.工业金属管道设计规范:GB 50316—2000[S].北京:中国计划出版社,2008:74-81.
- [8] 国家市场监督管理总局国家标准化委员会.压力管道规范工业管道:第3部分设计和计算:GB/T20801.3—2020[S].北京:中国标准出版社,2020:100-101.
- [9] 唐永进.压力管道应力分析:第2版[M].北京:中国石化出版社,2009:77-78.

(上接第39页)

- [3] 刘逸涵.基于电能表需求预测的配送优化研究[D].北京:北京交通大学,2017.
- [4] 陈剑强,杨俊杰,楼志斌.基于XGBoost算法的新型短期负

荷预测模型研究[J].电测与仪表,2019,56(21):23-29.

- [5] 李树卿,陈鼎,仇群辉,等.基于随机森林的电能质量综合评估[J].现代电力,2019,36(2):81-87.

(上接第43页)

- [4] 王志刚,张磊,许峰.基于机器学习的电力设备异常检测与分类方法[J].电力系统自动化,2017,41(6):130-137.
- [5] 张勇,李思,王鹏.电力设备绝缘监测与预警技术研究[J].高电压技术,2016,42(5):1349-1355.
- [6] 陈小平,刘振宇,张涛.电缆线路绝缘在线监测技术研究与应用[J].高电压技术,2015,41(9):2639-2645.
- [7] 李明,张华,王大力.电力设备过电压防护装置寿命评估方法研究[J].高电压技术,2017,43(4):978-985.

- [8] 周泽民,黄进,刘建华.基于光纤传感技术的电力设备温度监测方法研究与应用[J].光电工程,2016,44(8):196-202.
- [9] 张晓宁,李娜,王瑞.基于红外热像技术的电力设备温度异常检测方法研究[J].红外与激光工程,2018,47(6):134-139.
- [10] 李小云,张洪涛,王晓明.基于大数据技术的电力设备故障预测方法研究[J].电网技术,2019,43(1):233-240.
- [11] 孙江平.智能型剩余电流监测技术在变电站中的应用研究[J].电力与能源,2021,42(2):178-181.