

吴林峰, 陈源萍, 谭元刚. 基于细化用户画像的居民冷暖用电数据预测方法[J]. 智能计算机与应用, 2024, 14(5): 265-269.
DOI: 10.20169/j.issn.2095-2163.240538

基于细化用户画像的居民冷暖用电数据预测方法

吴林峰¹, 陈源萍², 谭元刚¹

(1 国网重庆市电力公司营销服务中心, 重庆 401100; 2 重庆广汇供电服务有限公司客户服务分公司, 重庆 401100)

摘要: 为提升电网运营数据应用价值与客户服务精益化管理水平, 针对温度变化影响居民用电负荷问题, 提出了一种基于细化用户画像+BP神经网络算法的居民冷暖用电数据预测方法。该方法首先利用时序分解对居民随温度变化的日用电量指标进行提取, 并基于此细化居民用电行为画像类, 随后基于细化类采用BP神经网络对居民冷暖电量进行预测。实验表明, 该算法可提升居民冷暖用电量预测准确度, 实现了对居民用电隐性动态特征的深度挖掘。

关键词: 时序分解; 模糊C均值聚类; BP神经网络; 冷暖用电预测; 细化用户画像

中图分类号: TM933

文献标志码: A

文章编号: 2095-2163(2024)05-0265-05

Prediction method of electricity consumption changing with temperature based on refined user portrait

WU Linfeng¹, CHEN Yuanping², TAN Yuangang¹

(1 Marketing Service Center, State Grid Chongqing Electric Power Company, Chongqing 401100, China;

2 Customer Service Branch of Chongqing Guanghui Power Supply Service Co., Ltd., Chongqing 401100, China)

Abstract: In order to improve the application value of power grid operation data and the lean management level of customer service, and aiming at the problem that temperature changes affect residents' electricity load, a prediction method for residents' heating and cooling electricity data based on refined user portrait + BP neural network algorithm is proposed. This method firstly uses time series decomposition to extract the daily electricity consumption index of residents that changes with temperature, and based on this, refines the portrait class of residents' electricity consumption behavior, and then uses BP neural network to predict the heating and cooling power of residents based on the refined class. Experiments show that the algorithm can improve the prediction accuracy of residential heating and cooling electricity consumption, and realize the deep mining of the hidden dynamic characteristics of residential electricity consumption.

Key words: time sequence decomposition; fuzzy C-means clustering; back propagation neural network; electricity consumption prediction changing with temperature; refining user portraits

0 引言

近年来, 国内科技和经济已然进入高质量发展阶段, 电网负荷调控^[1]也正逐渐向智能化管理阶段迈进, 随着新型电力系统的推进, 对电力负荷平衡提出了更高要求^[2]。了解海量用户的用电需求是电力公司和供电企业提升服务管理水平、增强竞争力的关键。尤其随着电能替代的推进, 负荷峰谷差进一步加大, 亟需对需求侧资源进行深入研究和分析。用户精细化特性分析, 即对用户画像进行多方位的刻画, 是获得用户用电需求的前提条件^[3]。且用户

行为画像的构建能够使用户隐性特征显性化, 辅助供电企业洞察用户需求、挖掘负荷调控潜力, 从而提升营销策略和电网辅助服务水平。

随着负荷调控工作的开展, 对用户用电行为分析进行了广泛研究^[4]。而在用户用电过程中, 用电量的影响因素较多, 国内外学者也从不同方面对此陆续发表了多项研究成果。文献[5]基于年度/月度负荷曲线, 对用户分段平均电量、波动率等特征进行提取, 从而分析用户的年度用电规律。文献[6]以削峰填谷为目标, 提取并分析用户用电高峰及季节性变化的特征量, 研究各类用户适合参与的需求

作者简介: 吴林峰(1989-), 男, 中级工程师, 主要研究方向: 电力营销相关研究; 陈源萍(1987-), 女, 会计, 主要研究方向: 电力营销; 谭元刚(1986-), 男, 高级工程师, 主要研究方向: 电力营销。

收稿日期: 2023-04-18

响应项目。文献[7]以年度用电量预测为主要分析目标,基于经济、气候及电价等关联影响因素对用户用电模式聚类分析方法进行了研究。

虽然上述方法均在一定程度上对用户用电行为进行刻画及分析,但缺乏对低压居民用户用电行为的分析,而研究表明温度因素直接影响低压居民的用电行为。故本文针对温度变化对居民用电量的影响,提出了一种基于细化用户画像+BP神经网络算法的居民冷暖用电量预测方法,以提升对低压居民

用电量的预测准确度,从而为电力系统的负荷资源调配提供数据支撑^[8]。

1 整体过程描述

本文基于历史居民日用电量数据,采用时序分解提取用户画像标签并细化用户类别,以此构建居民冷暖用电量预测模型,实现对居民冷暖用电量的预测。模型整体流程如图1所示。

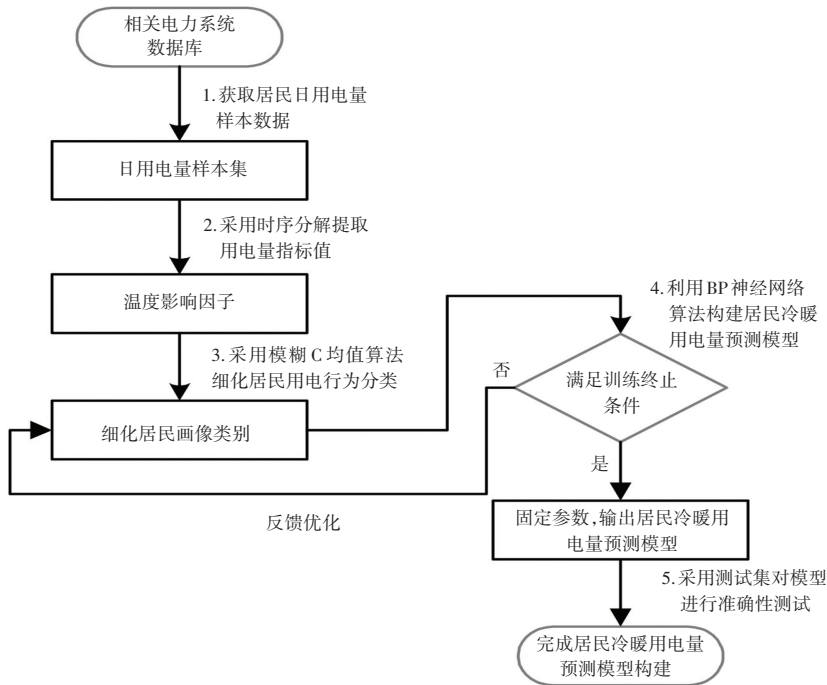


图1 整体流程图

Fig. 1 Overall flow chart

本文从营销业务应用系统数据库中获取历史居民日用电量数据,采用时间序列分解算法提取居民每日随温度变化的日用电量指标值,从而细化居民每日随温度变化的用电规律特性;随后,在细化用户画像分类基础上,在类别内根据时序分解后获取的温度影响因素,采用模糊C均值聚类算法(fuzzy C-means, FCM)进一步细化用户类别,提升类内用户用电量随温度变化规律的相似度;接着,基于不同的细化类别,采用BP神经网络(Back Propagation, BP)^[9]进行居民冷暖用电量的预测模型的构建,提升对居民日用电量的预测精度。

2 居民冷暖用电量分析及预测

2.1 采用时序分解提取温度影响因素

本文基于历史居民日用电量数据,采用时间序

列分解算法提取居民每日随温度变化的日用电量指标值。其中,定义每日温度特征包含日温度均值和日温差,代表不同温度时段。通过计算不同温度时段的用电量指标值,作为不同用户的温度影响因素。

基于获取的居民日用电量样本数据和日常温度时段数据,对连续时间的相同温度时段样本数据求用电量均值,随后计算日用电量的温度影响因素 P ,求解过程如下:

$$M_l = \frac{B_l + B_{l+1} + \dots + B_{l+N-1}}{N} \quad (1)$$

$$P_l = B_l / M_l \quad (2)$$

其中,若连续数天的温度时段相同,则将数天合并作为一个周期,并取其用户用电量均值为该周天的日用电量, N 表示周期数量, $N < 365$; B_l 表示用户在第 l 周期的用电量数据。

随后计算温度指数 r , 相当于计算以日为周期的移动平均值, 并从近 3 年的样本数据中挖掘随温度影响因子变化的用户日用电量变化规律:

$$r_l = \frac{\sum_{q=1}^T P_l^{(q)}}{T}, l = 1, 2, \dots, N \quad (3)$$

其中, T 表示获取样本的年周期数; $P_l^{(q)}$ 相当于第 q 年的第 l 个温度时段的温度影响因子值。

对 r 进行修正, 定义修正后的温度影响因子指数为 R , 解值过程为:

$$R_l = r_l / \left(\frac{\sum_{l=1}^N r_l}{N} \right) \quad (4)$$

根据上述过程获得用户在各个温度影响因子下的用电量变化影响指标。

2.2 模糊 C 均值聚类细化用户画像

模糊 C 均值聚类算法通过计算各待分类样本与聚类中心的隶属度, 实现样本的划分。算法以各簇类内距离最小、类间距离最大为优化目标, 动态调整样本的类属, 使聚类结果最优化。

基于已有的居民用电数据综合分析基础, 将居民用户行为分为以下几种类型: 工作日上班族、工作日居家族、休息日宅家族、休息日外出族、老人居家族、上班及老人混居族、空置型^[10]。在此基础上, 基于时序分解得到的各居民用户的温度影响因子值, 采用模糊 C 均值算法进行用户相似度聚类, 实现更加精准的相似居民用户分类。

给定待聚类居民用户用电数据样本集 $C = \{c_1, c_2, \dots, c_n\}$, $m_j = \{j = 1, 2, \dots, k\}$ 表示每个聚类的中心, k 表示类别数目, $\mu_j(c_i)$ 表示第 i 个样本对应第 j 类的隶属度函数, 则基于隶属度函数的聚类损失函数可以记为:

$$J = \sum_{j=1}^k \sum_{i=1}^n [\mu_j(c_i)]^b \|c_i - m_j\|^2 \cdot R_i \quad (5)$$

其中, $\sum_j \mu_j(c_i) = 1$, b 表示加权指数, 取值为 2。

令聚类损失函数 J 对 m_j 和 $\mu_j(c_i)$ 的偏导为 0, 求得式(5)的极小值的必要条件如下:

$$m_j = \frac{\sum_{i=1}^n [\mu_j(c_i)]^b c_i}{\sum_{i=1}^n [\mu_j(c_i)]^b} \quad (6)$$

$$\mu_j(c_i) = \frac{\|c_i - m_j\|^{-2/(b-1)}}{\sum_{s=1}^k \|c_i - m_s\|^{-2/(b-1)}} \quad (7)$$

将模糊 C 均值聚类的过程归纳为如下步骤:

(1) 输入待聚类样本集 C 和各居民用户的温度影响因子 R , 并初始化 $n * k$ 的隶属度矩阵 $U(0)$, $\mu_j(c_i)$ 为其中的第 i 行第 j 列的隶属度;

(2) 根据式(6)不断更新聚类中心 $m_j = \{j = 1, 2, \dots, k\}$;

(3) 根据式(7)更新隶属矩阵 $U(\mu_j(c_i))$;

(4) 通过对聚类中心和隶属度矩阵的不断更新迭代, 若满足终止条件 $\max_{ij} \{ \|(\mu_j(c_i))' - \mu_j(c_i)\| \} < \varepsilon$ 时, 则停止迭代, 输出细化居民用户类别; 否则返回步骤(2), 继续重复上述过程。

以工作日上班族为例, 经过聚类算法的分类后, 可输出温度敏感型、温度不敏感型和平稳型三种分类结果, 从而细化随温度变化的居民用户用电分类画像。

2.3 构建居民冷暖用电量预测模型

基于聚类获取到的细分居民用户画像类别, 采用 BP 神经网络算法进行居民冷暖用电量预测模型的构建, 实现对居民冷暖用电量的预测。

首先, 采用最大最小值标准化算法, 对用户日用电量进行标准化处理, 并提取其最大值、最小值和中间值, 用于后续预测用电量数据还原。随后, 采用 BP 神经网络算法构建居民冷暖用电量预测模型, 将居民用户的第 t 日用电量、第 t 日日常温度、第 t 日节假日标记、第 $t+1$ 日日常温度、第 $t+1$ 日节假日标记数据作为网络输入层的输入数据, 输出层输出居民冷暖用电量第 $t+1$ 日的预测用电量数据预测。数据处理过程为:

对样本数据进行标准化处理, 将最小值设定为 0, 最大值设定为 1, 中间值用式(8)来计算:

$$x_z = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (8)$$

其中, x_{\max} 、 x_{\min} 分别表示训练样本的最大值和最小值。

初始化输入层与隐含层连接权值 ω_{ph} , 输出层与隐含层连接权重 ω_{hg} , 隐含层阈值 α_h , 输出层的阈值 θ_g ; 设定输入层居民用户用电样本数据为 $X = (x_1, x_2, \dots, x_\eta)$, 期望输出为预测用电量数据 Y 。

计算隐含层输出数据, 隐含层输出 H 的计算为:

$$H_h = f\left(\sum_{p=1}^{\eta} \omega_{ph} \cdot d_p + \alpha_h\right), h = 1, 2, \dots, \zeta \quad (9)$$

其中, f 表示隐含层激励函数; p 表示输入层的节点数; h 表示隐含层的节点数; η 表示样本个数。

计算输出层的输出数据, H 为隐含层的输出, 则预测输出概率值 y 为:

$$y_g = \sum_{h=1}^{\xi} H_h \omega_{hg} + \theta_g, \quad g = 1, 2, \dots, \beta \quad (10)$$

其中, y_g 表示预测概率值, 将用户最大用电量 W 与 y_g 相乘即可获得预测的用电量值 Y 。

基于预测值与真实值误差, 误差函数 E 的计算为:

$$E_g = Y_g - O_g, \quad g = 1, 2, \dots, \beta \quad (11)$$

其中, O_g 表示实际居民冷暖用电量值。

根据预测误差 E 更新网络的连接权值和阈值:

$$\omega_{ph} = \omega_{ph} + \psi H_h (1 - H_h) d_p \sum_{g=1}^{\beta} \omega_{hg} E_g \quad (12)$$

$$\omega_{hg} = \omega_{hg} + \psi H_h E_g \quad (13)$$

$$\alpha_h = \alpha_h + \psi H_h (1 - H_h) \sum_{g=1}^{\beta} \omega_{hg} E_g \quad (14)$$

$$\theta_g = \theta_g + \psi E_g \quad (15)$$

其中, ψ 表示权重参数。

上述过程不断反馈迭代, 以修正网络连接权值和阈值, 直至预测误差满足预定要求后, 便可以停止模型训练, 固定参数, 完成居民冷暖用电量预测模型的构建, 用以实现对居民用电量的预测。

3 实验结果与分析

本文以国网某省营销业务应用系统为实验背景平台, 依照细化聚类后的居民用户画像类型, 选取某一居民小区的用电数据为例构建居民冷暖用电量预测模型。选取该小区 962 个用户 3 年来的日用电量数据作为训练样本集, 剔除异常样本后, 共计获取 350 例样本集, 128 例为正常样本; 在该小区居民用户中, 选取 65 个用户日用电量数据样本作为测试样本, 对居民冷暖用电量进行预测。从营销业务应用系统获取居民档案类数据、用电信息采集系统获取居民日用电量数据、基于网页爬取获取每日温度数据和节假日数据, 收集上述数据作为样本特征数据集。

3.1 模型预测准确率对比

本实验部分选取测试样本集的数据对该模型的有效性进行测试。同时, 将本文方法与采用 BP 神经网络、指数平滑和 GRU 三种算法构建的居民用电量预测模型进行预测准确率对比。

(1) BP 神经网络: 无细化用户画像类别, 以传统的用户分类为数据样本集基础, 采用 BP 神经网络构建居民用电量预测模型;

(2) 指数平滑: 是一种特殊的加权移动平均法, 通过选择不同的平滑常数来调节时间序列趋势变化的平稳程度, 从而实现对居民用电量的预测;

(3) GRU: 循环神经网络的一种, 基于长期短时

记忆和反向传播调参构建居民用电量预测模型。

在测试中, 采用平均绝对误差 (Mean Absolute Error, MAE) 作为指标来衡量居民冷暖用电量预测的准确性, 定义如下:

$$E_{MAE} = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (16)$$

其中, Y_i 表示用电量预测值, \hat{Y}_i 表示用电量真实值。

本文所提方法与采用其他 3 种算法构建居民用电量预测模型, 准确性对比结果见表 1。

表 1 预测准确性对比

方法名称	MAE	%
本文方法	92.89	
BP 神经网络	82.17	
指数平滑	83.46	
GRU	86.35	

此实验对比说明采用其他 3 种算法对居民冷暖用电量预测的准确率。由对比结果可以看到, 其他算法虽也具有不错的预测性能, 但是本文所提方法的预测准确率展示出明显的优势。实验结果表明, 本文采用细化用户画像+BP 神经网络算法构建的居民冷暖用电量预测模型, 有效提升了居民冷暖用电量的预测准确率。

3.2 居民冷暖用电量预测分析

基于本文构建的居民冷暖用电量预测模型, 分别对测试样本集中的 65 户居民每天的冷暖用电量进行预测, 并以月为单位, 统计和计算 65 户居民预测用电量的累计值, 预测的时间范围为 2021 年 3~7 月。居民用户的真实用电量数据可从营销业务系统中直接获取。图 2 展示了真实用电量与预测用电量的对比情况。

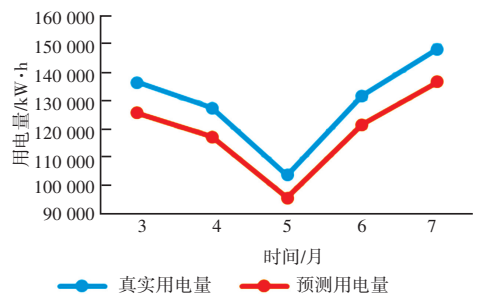


图 2 用电量真实值与预测值对比图

Fig. 2 Comparison chart of actual and predicted electricity consumption

对图2进行纵向分析可以看出,用电量预测值与真实值相比,误差在10%以内,表明本文所构建的居民冷暖用电量预测模型具有良好的性能。对图2进行横向对比可以看出,温度的变化在一定程度上对居民用电量产生影响。在3月份,温度较低,故而居民需进行多种方式的用电取暖,用电量随之变高;在5月份,温度适宜,既不需要降温、也不需要取暖,所以用电量达到全年最低;在7月份,进入夏季,温度偏高,居民陆续开始防暑降温,用电量再次升高。本文方法可有效捕获居民用电量随温度的变化规律,可为国网公司开展负荷调控及供需平衡等工作提供数据支撑。

4 结束语

本文提出了一种基于细化用户画像+BP神经网络算法的居民冷暖用电数据预测方法。该方法首先利用时序分解细化居民用户每日随温度变化的用电规律特征;随后,采用模糊C均值聚类算法进行用户细化分类,提升类内用户用电量随温度变化趋势的相似度;接着,基于不同的画像分类,构建基于BP神经网络的居民冷暖用电量预测模型,完成对居民用户行为画像的精准刻画,有助于提升业务数据分析能力以及服务精益化管理水平^[11]。

参考文献

- [1] 马杰,李秋燕,丁岩,等. 含高渗透率可再生能源的配电网灵活性评价指标体系及计算方法[J]. 电力系统及其自动化学报, 2020, 32(9): 99-104, 112.
- [2] 王永明,陈宇星,殷自力,等. 基于大数据分析的电力用户行为画像构建方法研究[J]. 高压电器, 2022, 58(10): 173-179, 187.
- [3] 唐辉,刘晓波,韩祥民,等. 基于特征加权的SFO-K_means用电行为研究[J]. 智能计算机与应用, 2021, 11(9): 165-169.
- [4] 王磊,刘洋,李文峰,等. 基于用电行为数字特征画像的电力用户两阶段分类方法[J]. 电力建设, 2022, 43(2): 70-80.
- [5] 孙毅,贾孟扬,陆俊,等. 计及用户需求响应的智能用电互动潜力分析[J]. 电力科学与技术学报, 2016, 31(4): 43-50.
- [6] 苏适,李康平,严玉廷,等. 基于密度空间聚类和引力搜索算法的居民负荷用电模式分类模型[J]. 电力自动化设备, 2018, 38(1): 129-136.
- [7] 赵腾,王林童,张焰,等. 采用互信息与随机森林算法的用户用电关联因素辨识及用电量预测方法[J]. 中国电机工程学报, 2016, 36(3): 604-614.
- [8] 丘嘉苑,汤平瑜,穆青青. 基于多源数据处理的用户用电行为侧写技术[J]. 供用电, 2022, 39(2): 80-85.
- [9] 苏颜,张珍,林庆达,等. 基于BP神经网络算法的短期电力负荷预测研究[J]. 电子设计工程, 2022, 30(12): 167-170, 175.
- [10] 徐涛,黄莉,李敏蕾,等. 基于多维细粒度行为数据的居民用户画像方法研究[J]. 电力需求侧管理, 2019, 21(3): 47-52, 58.
- [11] 黄元行,刘晓波,范津玮,等. 基于注意力机制和EMD-GRU模型的电力负荷预测[J]. 智能计算机与应用, 2021, 11(3): 33-37, 43.