

# 基于属性挖掘的供电所用户用电行为画像模型

王海燕, 薄博, 曹敬立, 李思源, 李磊, 梁佳宁

(北京工业大学 计算机科学与技术学院, 河北 承德 067000)

**摘要:** 由于已有模型未能在构建用户行为画像的过程中加入属性挖掘, 导致供电所用户用电行为分析结果不理想, 模型建立时间较长, 因此构建提出基于属性挖掘的供电所用户用电行为画像模型。采用主成分分析方法对负荷数据降维处理; 通过属性挖掘对降维数据聚类分析, 区分不同用户的用电行为属性。针对供电所各个类别的用电趋势曲线进行建模分析, 同时结合特征标签体系和总体类别对用户用电行为进行可视化呈现。仿真实验结果表明, 所提模型能够获取更加准确的供电所用户用电行为分析结果, 同时还能够有效减少模型的建立时间。

**关键词:** 属性挖掘; 供电所; 用户用电行为; 画像模型

中图分类号: TP18 文献标识码: A 文章编号: 1003-7241(2025)05-0178-04

## Portrait Model of Power Consumption Behavior of Power Supply Station Users Based on Attribute Mining

WANG Hai-yan, BO Bo, CAO Jing-li, LI Si-yuan, LI Lei, LIANG Jia-ning

(Beijing University of Technology, Information Technology, Chengde 067000 China)

**Abstract:** Since the existing model is not able to add attribute excavation in the process of building user behavior portraits, it is not satisfactory that the user's electrical behavior analysis results is not ideal. The model has a long time, so it is constructed to propose an electrical behavior portrait of a user based on attribute excavation model. Using the main component analysis method to reduce load data designation, it analysis of the dodensaceous data clustering analysis of different users by attribute excavation, distinguishes the electrical behavior attributes of different users. Modification analysis of the electricity trend curve of the various categories of power supply, while combining the feature label system and the overall category to visualize the user's electric behavior. Simulation experiment results show that the model can obtain more accurate power supply user's electrical behavior analysis results, and can also effectively reduce the establishment time of the model.

**Keywords:** attribute mining; power supply station; user electricity consumption behavior; portrait model

## 0 引言

自从进入20世纪以后, 科学技术的飞速发展和可持续能源的日益重视, 智能电网概念被提出<sup>[1-2]</sup>。在过去的几十年里, 智能电网在全球相关政策的支持下得到飞速发展。随着人类积累的用电数据日益增加<sup>[3]</sup>, 如何在海量数据中提取有价值的信息成为研究的主要内容。目前, 电力系统的用户用电行为画像模型主要利用数据挖掘的形式完成对数据的学习和挖掘, 但是通过查阅相关资料发现, 仍然还存在部分问题未能得到有效解决。为了满足时代发展需求, 国内外相关专家在上述模型的基础上进行更加深入的研究。例如赵晋泉等人<sup>[4]</sup>优先根据组建聚合回报指标, 同时兼顾聚集度和分析, 完成用户分类; 然后通过k-means方法将最大相关最小冗余准则应用于用户用电行为特征选择中; 最优利用遍历法获取最优特

征子集, 对全部优质特征进行量化处理, 有效实现用户画像构建。董哲瑾等人<sup>[5]</sup>利用散点图分析影响用户成长值的因素, 提取行为特征和相对稳定的时间特征, 同时对全部特征进行筛选, 完成用户画像构建。以上两种模型由于未能在构建的过程中加入属性挖掘方法, 导致供电所用户用电行为分析结果和真实结果存在较大误差, 模型建立时间增加。为此, 提出一种基于属性挖掘的供电所用户用电行为画像模型。仿真实验结果表明, 所提模型能够快速准确完成供电所用户用电行为画像模型建立。

## 1 供电所用户用电行为画像模型

### 1.1 基于属性挖掘的供电所用户用电行为挖掘

基于属性挖掘的供电所用户用电行为划分操作步骤如图1所示, 主要包含预处理以及聚类等操作步骤。其中输入为供电所内部采集系统所采集的负荷数据, 输出为用户用电行为档案。

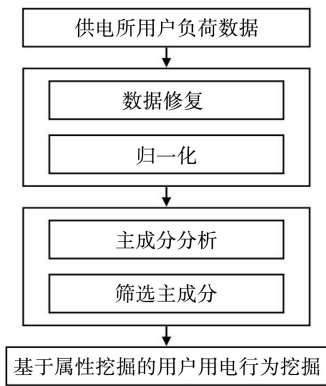


图1 基于属性挖掘的供电所用户用电行为划分流程图

(1) 由于供电所所在环境的不同,促使负荷环境的准确性会受到不同程度的影响。所以,采集到的数据会出现数据缺失或者损坏等情况的发生。另外,这些数据的存在也会影响用户用电行为的划分。为了解决上述问题,需要对缺失或者损坏的数据进行修复和完善,同时求解平均值,具体的计算式如下所示:

$$x_{i,t}^y = \frac{x_{i,t}^{y-3} + x_{i,t}^{y-2} + x_{i,t}^{y+1} + x_{i,t}^{y-1} + x_{i,t}^{y+2} + x_{i,t}^{y+3}}{6} \quad (1)$$

式中,  $x_{i,t}^y$  代表用户  $i$  在第  $y$  天第  $t$  个计量时刻的负荷数据;  $x_{i,t}^{y-3}$ 、 $x_{i,t}^{y-2}$ 、 $x_{i,t}^{y+1}$ 、 $x_{i,t}^{y-1}$ 、 $x_{i,t}^{y+2}$  和  $x_{i,t}^{y+3}$  分别代表用户  $i$  在第  $y-3$  天、 $y-2$ 、 $y+1$ 、 $y-1$ 、 $y+2$  和第  $y+3$  天第  $t$  个计量时间的负荷数据。

为了更好消除功率数据量级产生的影响,需要更加精准描述不同数据集中用户的用电趋势,同时对采集到的数据进行归一化处理,具体的操作过程如公式(2)所示。

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (2)$$

式中,  $x$  代表归一化处理的负荷数据;  $x_{\max}$  和  $x_{\min}$  分别代表负荷数据的最大值和最小值。

## (2) 降维

为了有效降低基于属性挖掘的供电所用户用电行为挖掘所使用的时间,需要优先对各个类型的负荷数据进行降维处理,以下主要通过PCA算法对采集的电力负荷数据进行降维处理。

设定供电所共有  $n$  个电力用户,各个用户共包含  $m$  个采集的负荷数据,则构建的负荷数据矩阵  $X$  能够表示为:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \cdots & \cdots & \cdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} = (X_1, X_2, \cdots, X_n) \quad (3)$$

对于  $(X_1, X_2, \cdots, X_n)$  而言,需要通过特征值分解协方差矩阵完成PCA算法点的降维操作,具体操作步骤如下所示:

1) 输入电力负荷矩阵  $(X_1, X_2, \cdots, X_n)$ , 其中  $X_1, X_2, \cdots, X_n$  均为  $m$  维;

2) 计算平均取值,如公式(4)所示;

$$\tilde{x} = \frac{1}{n} \sum_{i=1}^m x_i \quad (4)$$

式中,  $\tilde{x}$  代表平均值;

3) 计算协方差矩阵  $\frac{1}{n} XX^T$ ;

4) 计算协方差矩阵的特征值和对应的向量;

5) 将特征值按照从大到小的顺序进行排列,选取最大的  $k$  个向量构建特征向量矩阵。

(3) 基于属性挖掘的供电所用户用电行为挖掘

为了精准区分不同用户的用电趋势,通过属性挖掘对用户用电行为进行挖掘,有效区分不同用户的用电属性,以下主要从初始聚类数量或者初始聚类中心两个方面进行分析研究,进而完成用户用电行为挖掘:

1) 对初始聚类数目进行计算

全面分析初始聚类数量对  $k$ -means 算法聚类效果产生的影响<sup>[6]</sup>,通过公式(5)确定最佳的  $k$  取值:

$$k = \arg \max (L_k(v)) \quad (5)$$

式中,  $v$  代表聚类总数。

其中:

$$L_k(v) = \frac{D_{IEF}(v)}{D_{IEF}(v+1)} \quad (6)$$

$$D_{IEF}(v) = (v-1)^{\frac{2}{n}} D_{v-1} - v^{\frac{2}{n}} D_v \quad (7)$$

式中,  $n$  代表聚类的维数;  $D_v$  代表聚类距离的平方和;  $D_{IEF}$  代表簇内聚类的变化规律;

2) 计算初始聚类中心

通过公式(8)确定各个簇的聚类中心,针对采集到的负荷数据曲线集合,设定负荷曲线的目的参数。将  $X'_y$  设定为初始聚类中心曲线,其中负荷数据曲线集合内全部的负荷曲线和  $X'_y$  的距离小于平均距离的数量,将其表示为  $D_{density}(X'_y, MeanDist)$ , 具体的计算式为:

$$D_{density}(X'_y, MeanDist) = \sum_{i=1}^n u \left( MeanDist - dist(X_i - X'_y) \right) \quad (8)$$

其中:

$$\begin{cases} MeanDist = \frac{1}{c_n^2} \sum_i d(X_i, X_j) \\ d(X_i, X_j) = \sqrt{\sum_{q=1}^m (X_{iq} - X_{jq})^2} \end{cases} \quad (9)$$

式中,  $d(X_i, X_j)$  代表欧式距离;  $MeanDist$  代表最短聚类中心;

3) 计算各个负荷曲线到质心的欧氏距离,选取最短的聚类质心到达负荷曲线的簇集合;

4) 重复上述操作过程,直至质心不再发生任何变化。

(4) 输入最终的用户用电行为挖掘结果

将通过  $k$ -means 算法对用户用电属性的聚类结果设定为最终的用户属性挖掘结果。

### 1.2 供电所用户用电行为画像模型的建立

通过k-means算法对用户用电行为进行挖掘后,结构挖掘结果分别组建不同的用户类别,为后续供电所用户用电行为画像模型的建立奠定坚实的基础。

将用户的总体用电特征作为目标需求导向,同时针对结合挖掘结果,进行分段分析,具体的分析步骤如下所示:

#### (1) 数据标准化处理

针对相同时段的用户用电行为特征<sup>[7]</sup>,按照类别进行划分,其中供电所的用户数据样本能够表示为:

$$L = \begin{bmatrix} L_1(1), L_1(2), L_1(3), \dots, L_1(k), \dots, L_1(14) \\ L_2(1), L_2(2), L_2(3), \dots, L_2(k), \dots, L_2(14) \\ \vdots \\ L_n(1), L_n(2), L_n(3), \dots, L_n(k), \dots, L_n(14) \end{bmatrix} \quad (10)$$

其中,14维数据分别代表不同的物理意义,整体量值范围差异变化趋势较为明显,同时还存在大量的异常数据。在进行综合分析的过程中,优先对数据进行标准化处理,则有:

$$L(k) = \left( L(k) - \frac{\bar{L}(k)}{\sigma(L(k))} \right) \quad (11)$$

式中, $\bar{L}(k)$ 代表 $L(k)$ 的平均取值; $\sigma(L(k))$ 代表 $L(k)$ 的方差。

#### (2) 对各项指标进行综合分析

##### 1) 用户行为特征标签 $L_B$

将用户行为特征标签 $L_B$ 设定为逻辑量,分别放置到对应的调控目标中,同时将行为标签设定为用户大群初选,将无价值的用户过滤掉。

##### 2) 用电特性特征标签 $L_E$

用电特性特征标签主要包含8种不同类型的数据,通过公式(12)求解不同用户用电特性的综合标签 $L_E$ ,即:

$$L_E = \sum_{k=2}^9 \lambda_k L(k) \quad (12)$$

式中, $\lambda_k$ 代表不同维数的数据。

##### 3) 消费习惯特征标签 $L_C$

主要根据不同用户的用电特性,组建消费习惯特征标签 $L_C$ 。

#### (3) 供电所用户用电行为分析

主要利用轮廓系数CH指标确定最优类数,经过步骤(2)之后将数据集转换为二维数据 $(L_E(i), L_C(i))$ 。同时针对用户的用电特性标签 $L_E(i)$ ,获取不同类别的中心点 $L_E(k)$ ,即:

$$\min \sum_{k=1}^{m1} \sum_{i=1}^n (L_E(i) - L_E(k))^2 \quad (13)$$

通过 $L_E$ 和 $L_C$ 中心点进行组合 $(L_E(k), L_C(i))$ ,构成 $m1 \times m2$ 个类别。针对全面中心点,主要利用曼哈顿距离对各个用户的类别进行判定<sup>[8]</sup>,具体的计算式为:

$$\min (abs(L_E(i) - L_E(k))) = abs(L_C(i) - L_C(j)) \quad (14)$$

#### (4) 用户类别分析

将 $L_E$ 和 $L_C$ 划分为三种不同的类型,同时以客观性指标为判定原则,将类别进行纵向排序,同时依次增加,具体如图2所示。

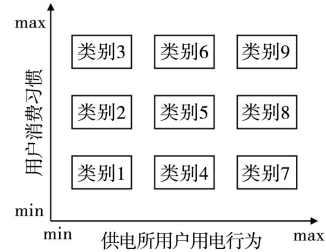


图2 供电所用户调控级别分布图

将供电所的用户用电行为标签特征值和总体调控类别两者相结合,构建供电所用户用电行为画像模型,同时通过模型直观展示用户的用电行为。

## 2 仿真实验分析

为了验证所提基于属性挖掘的供电所用户用电行为画像模型的有效性,选取H城市作为研究对象,重点针对城市居民的用户用电行为画像进行分析。

#### (1) 供电所用户用电行为画像模型分析结果

主要通过供电所用户用电日负荷曲线变化和平均绝对比百分误差衡量模型分析结果的好坏,具体实验结果如下所示:

选取H城市典型的用户作为研究对象,分别采用三种不同模型对用户的用电日负荷变化曲线进行分析研究,具体实验结果如图3所示。

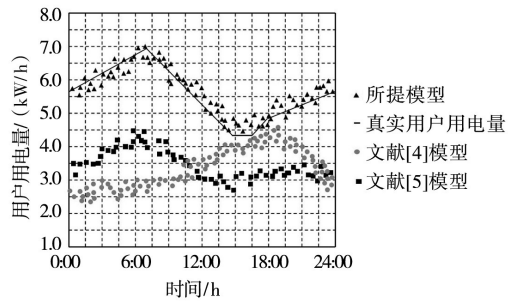


图3 不同模型的供电所用户用电日负荷曲线变化情况

分析图3中的实验数据可知,用户的用电日负荷变化曲线在不同时间段均不同。实验主要选取三种不同的模型对用户的用电日负荷曲线变化情况进行分析,其中,所提模型的分析结果和用户的真实用电量更加接近,全面验证了所提模型的优越性。

为了更进一步验证所提模型的优越性,以下实验通过平均绝对比百分误差衡量模型分析结果的准确性,具体实验结果如表1所示。

分析表1中的实验数据可知,其中所提模型分析结果

的平均绝对比百分误差在三种模型中为最低;剩余两种模型的平均绝对比百分误差明显更高一些。充分证明所提模型在建立过程中加入属性挖掘的可行性和有效性。

表1 不同模型的平均绝对比百分误差对比结果

用户类型	平均绝对比百分误差/%		
	所提模型	文献[4]模型	文献[5]模型
A1	0.05	0.06	0.07
A2	0.03	0.08	0.06
A3	0.02	0.05	0.07
A4	0.01	0.07	0.10
A5	0.05	0.09	0.08
A6	0.04	0.07	0.06
A7	0.02	0.03	0.08
A8	0.01	0.04	0.07
A9	0.03	0.06	0.02

(2) 供电所用户用电行为画像平均建立时间

由于不同类型的用户用电行为存在明显差异,以下实验分别针对各个类型的用户用电行为画像平均建立时间进行对比,实验结果如表2所示。

表2 不同模型的供电所用户用电行为画像平均建立时间对比结果

用户类型	供电所用户用电行为画像平均建立时间/ms		
	所提模型	文献[4]模型	文献[5]模型
A1	58	63	61
A2	50	55	57
A3	46	51	53
A4	42	47	59
A5	47	53	54
A6	44	49	56

分析表2中的实验数据客户,由于用户类型的差异,促使各个模型的平均建立时间存在十分较大的差距。由于所提模型在建立的过程中,对各个用户的行为属性进行挖掘和分析,简化画像建立过程,进而有效降低画像的建立时间。

(上接第112页)

[5] 王文悦,刘海涛,季宇.虚拟电厂可调空间统一建模及其参与调峰市场的优化运行策略[J].电力系统自动化,2022,46(18):74-82.

[6] 郑晓端.基于大数据的电网企业供电成本预测模型及其构建[J].企业改革与管理,2020(4):179-181.

[7] 熊一,詹智红,柯方超,等.基于改进BP神经网络的变电站检修运维成本预测[J].电力科学与技术学报,2021,36(4):44-52.

[8] 杨久红,王小增.基于改进BP神经网络的产品成本预测研究和实现[J].机械设计与制造,2010(7):223-225.

[9] 南敬昌,臧净,高明明.改进蚁群算法的BRBP神经网络放逆向建模方法[J].激光与光电子学进展,2020,57(1):198-205.

[10] 王云艳,罗帅,王子健.基于改进型BP神经网络的光伏功率预测[J].计算机仿真,2022,39(11):153-157.

[11] 林杰,贺键.大数据环境下基于PCA分层抽样的并行随机

3 结束语

针对模型存在的一系列问题,提出一种基于属性挖掘的供电所用户用电行为画像模型。仿真实验结果表明,所提模型的用户用电日负荷曲线和用户的真实用电量更加接近,同时还能够有效降低用户用电行为画像平均建立时间和分析结果平均绝对比百分误差,获取高准确性的用户用电行为画像。

未来将针对用户用电行为进行更加深入的研究,虽然取得了一些显著的成果,但是仍然存在以下几方面的问题需要进一步完善:(1)进一步对用户用电行为和外部影响因素进行分析。(2)后续研究的过程中加入深度学习算法,促使模型的建立速度进一步得到提升。

参考文献:

[1] 于兴尚,王迎胜.面向精准化服务的图书馆用户画像模型构建[J].图书情报工作,2019,635(22):42-49.

[2] 谭浩,郭雅婷.基于大数据的用户画像构建方法与运用[J].包装工程,2019,412(22):106-112.

[3] 冯秋燕,朱学芳.社交媒体用户价值画像建模与应用研究[J].情报资料工作,2019,231(6):75-82.

[4] 赵晋泉,夏雪,刘子文,徐春雷,苏大威,闪鑫.电力用户用电特征选择与行为画像[J].电网技术,2020,442(9):322-330.

[5] 董哲瑾,王健,钱凌飞,等.一种用户成长性画像的建模方法[J].山东大学学报(理学版),2019,54(3):42-49.

[6] 孙铁柱,田琳.基于CRT分类算法的用户画像分层模型——以银行借贷用户为例[J].情报科学,2020,349(9):77-83.

[7] 刘亚辉.多路隔离本安电源输出故障识别电路设计[J].能源与环保,2020,42(4):124-128.

[8] 李堃,黎向锋,杨振泰,等.基于夜间辅助驾驶的图像预处理方法研究[J].机械制造与自动化,2020,49(2):162-165.

作者简介:王海燕(1981-),女,硕士研究生,高级工程师,研究方向:农电管理。

森林预测算法[J].中国管理信息化,2022,25(24):172-176.

[12] 郭金玉,张安宝,李元.二阶差商PCA算法研究及应用[J].沈阳化工大学学报,2021,35(1):61-68.

[13] 张振莲,鲁淑霞,翟俊海.邻域欠采样的AdaBoost\_v算法[J].小型微型计算机系统,2023,44(1):97-102.

[14] 吕佳朋,史贤俊.基于AdaBoost的GPR预测算法研究及应用[J].光电与控制,2020,27(6):43-46,62.

[15] 王林峰,徐楠,聂婧,等.基于PCA-BP神经网络的输电线路工程投资估算模型研究[J].电气传动,2023,53(9):41-48.

作者简介:蔡晟琦(1985-),男,本科,工程师,研究方向:电网调度及电力市场运行,源网荷储多元协同控制技术。