

DOI:10.12158/j.2096-3203.2019.06.004

基于欠完备自编码器的用户用电行为分类分析方法

黄奇峰¹, 杨世海¹, 邓欣宇², 陈海文², 王守相²

(1. 国网江苏省电力有限公司电力科学研究院, 江苏 南京 211103;

2. 天津大学智能电网教育部重点实验室, 天津 300072)

摘要:针对电力大数据背景下用户用电行为复杂多变、分析困难的问题,提出了一种基于欠完备自编码器的用户用电行为分类分析方法。首先,通过欠完备自编码器对智能电表数据进行编码,实现对原始数据的特征抽取,并使用反向传播(BP)神经网络进行用户用电行为分类分析;然后,对最佳编码比率进行优选,并结合用户的典型用电特征作为神经网络的输入,提高了分类准确率;最后,在爱尔兰智能电表数据集上进行了仿真实验,并与直接使用BP神经网络进行对比,分析表明,文中所提出的用户用电行为分类分析方法不仅可以提高检测准确率,帮助电力公司更好地掌握用户用电规律,辅助需求响应实施,还能显著降低算法的运行时间。

关键词:欠完备自编码器;用户用电行为分析;需求响应;特征挖掘;智能用电;智能电表

中图分类号:TM73

文献标志码:A

文章编号:2096-3203(2019)06-0024-07

0 引言

近年来,随着智能电网的快速发展与信息化水平的不断提高,用户用电信息采集系统的建设趋于完善^[1]。智能电表是用电信息采集系统的基本量测设备之一,承担着电力数据的采集、计量和传输任务,同时具备简单的数据分析功能^[2-3]。智能电表的普及为电力公司带来海量用电数据,在此基础上对居民用电行为进行分类分析,掌握更加精细化的用户用电特征,充分挖掘需求响应潜力^[4-5]。

目前,国内外学者已对用户用电行为分类分析方法开展了很多研究。文献[6]提出了一种利用奇异值分解技术进行用户日负荷曲线分类的方法,具有运行时间短、鲁棒性好的优点。文献[7]利用类间优化与类内优化增强了负荷数据的可分性,且具备一定的自愈优化能力。文献[8]针对海量且分散性高的用户负荷数据,提出了一种分布式聚类算法,显著提高了负荷曲线聚类的有效性和鲁棒性。文献[9]基于模糊聚类算法与曲线相似度对负荷用户进行分类与识别,为电力公司精准服务提供依据。文献[10]提出了一种考虑负荷纵向随机性的基于经验模态分解的负荷聚类分析方法,从横向和纵向2个角度全面表征用户的用电行为。文献[11]提出一种基于信息熵分段聚合近似和谱聚类的负荷分类方法,获得了良好的负荷聚类效果。文献[12]利用用户用电模式数提高了短期负荷预测的准确度,为电力营销需求响应业务的快速发展提

收稿日期:2019-06-24;修回日期:2019-08-10

基金项目:国家重点研发计划资助项目“城区用户与电网供需友好互动系统”(2016YFB0901100)

供技术支撑。文献[13]提出了一种特征优选策略,在此基础上,文献[14]利用极限学习机对居民用电行为进行分类分析,并对网络参数进行优选,提高了检测的正确率。以上研究在用户用电行为分类分析方面取得了很好的效果,但并未充分考虑在特定应用场景中的适用性。

从实际应用的角度考虑,若能将用户用电行为分类分析功能集成在智能电表中,不仅可以充分利用智能电表的数据处理能力,还能有效提高调度人员的数据分析效率,降低人工成本^[15]。然而,受目前智能电表硬件条件的限制,高级功能集成需要考虑数据容量与计算效率。在保证算法准确率的前提下,研究降低数据量并提升计算效率的用户用电行为分类分析方法,是将该功能在智能电表中进行实际应用的必要条件。

针对上述需求,提出一种基于欠完备自编码器的用户用电行为分类分析方法。通过欠完备自编码器对原始电表数据进行特征抽取,使用反向传播(back propagation, BP)神经网络进行用户用电行为分类分析。在此基础上,对最佳编码比率进行优选,结合用户的典型用电特征,提升分类的准确率。经仿真验证,文中设计的用户用电行为分类分析方法不仅准确率高、分类误差小,且使用的数据量少,计算效率更高,具有较高的实际工程应用价值。

1 基于欠完备自编码器的用户用电行为分类模型

1.1 基本原理

自编码器(auto-encoder, AE)是一种由编码器

与解码器构成的 3 层神经网络无监督学习算法,经过中间隐层的非线性映射,将输入信号 X 编码得到一个新的信号 Y ,再解码回到 $X^{[16]}$ 。因此,自编码器的输入与输出形式几乎相同。若隐藏层的维数大于输入层,则得到过完备自编码器。若强制限制隐藏层的维度小于输入维度,则可得到欠完备自编码器 (undercomplete auto-encoder, UAE),学习欠完备的表示将强制自编码器捕捉训练数据中最显著的特征。UAE 工作原理如图 1 所示。

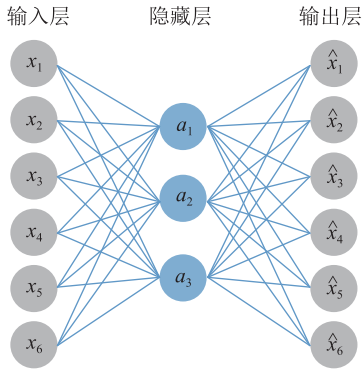


图 1 欠完备自编码器工作原理

Fig.1 Principle of undercomplete auto-encoder

对于给定的用户日用电数据 $X = \{x_1, x_2, \dots, x_N\}$,欠完备自编码器首先将 X 映射到中间隐藏层 (也称作特征层),学习 X 欠完备的表示,映射函数定义为:

$$h_i = s(wx_i + b) \quad (1)$$

式中: w 为编码权重系数; b 为编码偏执系数; h_i 为原始数据的编码结果; $s(x)$ 为编码器的激活函数。

学习欠完备的表示将强制欠完备自编码器捕捉智能电表数据中最显著的特征,即特征抽取,得到 $H = \{h_1, h_2, \dots, h_M\}$ (M 为经过编码得到的数据维度)。然后 H 被映射到输出层进行数据重构,映射函数定义为:

$$\hat{x}_i = s(w'h_i + b') \quad (2)$$

式中: w' 为解码权重系数; b' 为解码偏执系数; \hat{x}_i 为对 h_i 解码后的数据。

想要尽可能地对输入进行重构,则要求损失函数尽可能小,损失函数 $L(x, \hat{x})$ 定义为:

$$L(x, \hat{x}) = \frac{1}{N} \sum_{i=1}^N \|x_i - \hat{x}_i\|^2 \quad (3)$$

欠完备自编码可以很好地克服普通自编码器容易过拟合的缺陷,且中间隐藏层能够进行高效的特征抽取,但其存在中间隐藏层单元数特别少,导致重构过程比较困难的问题。当只使用网络的前半部分时,可以发挥欠完备自编码器的优势,实现对智能电表数据的降维。相比传统主成分分析 (principal component analysis, PCA) 每层之间的线性变换,欠完备自编码器采用的非线性变换可以学习到更重要、更全面的数据特征,将智能电表采集的高维原始数据降至能够保留重要特征的低维数据,有效降低数据的冗余度,从而减小数据所需的存储空间,提升算法的运行效率,以适应在智能电表上的应用场景^[17]。

文中将编码后的低维数据输入 BP 神经网络,并制定用户用电行为标签进行网络训练,实现对用户用电行为的分类,流程如图 2 所示。

1.2 编码比率优选

在使用欠完备自编码器进行编码时,编码比率 β 定义为:

$$\beta = \frac{M}{N} \quad (4)$$

式中: N 为原始电表数据维度; M 为编码得到的特征数据维度。 β 值越大,可以保留越多的数据特征,但所需的存储容量越大,降低了在智能电表上的适用性; β 值太小,又容易丢失重要的数据特征,降低了用户用电行为分类的准确率。只有选择合理的 β 值才能兼顾检测的正确率与数据占用的存储空间。

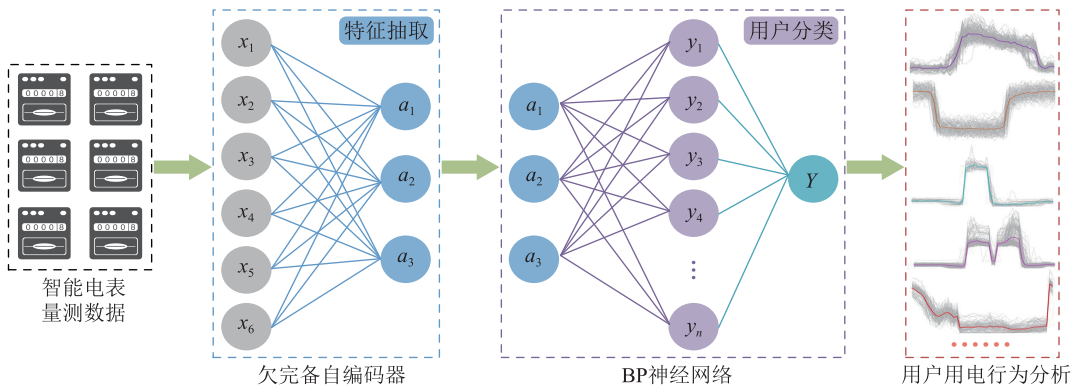


图 2 用户用电行为分类分析流程

Fig.2 Process of residential electricity consumption behavior classification analysis

F_1 指标是多分类问题中常用的评价指标, 兼顾分类的准确率和召回率, 其定义为:

$$F_1 = 2 \frac{P_r R_c}{P_r + R_c} \quad (5)$$

式中: P_r 为准确率, 其含义为预测为正的样本中, 真实为正的样本所占的比率; R_c 为召回率, 其含义为真实为正的样本中, 被预测为正的样本所占的比率。

文中选择 F_1 指标作为用户用电行为分类效果的评价指标, 并探究欠完备自编码器的最佳编码比率, 编码比率优选方案流程如图 3 所示。

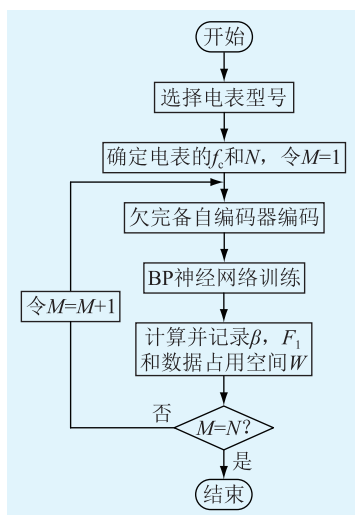


图 3 编码比率优选方案

Fig.3 Coding ratio optimization scheme

(1) 根据智能电表型号确定电表的数据采集频率 f_c , 计算日负荷数据点的维度 N , 同时设置欠完备自编码器中间隐藏层单元个数 $M=1$;

(2) 使用欠完备自编码器对电表原始数据进行编码, 获取中间隐藏层的特征数据 F , 将其作为 BP 神经网络的输入;

(3) 按照一定的标准制定用户用电行为标签, 进行网络训练, 计算测试集的 F_1 指标;

(4) 根据 M 和 N 计算 β 值, 记录 F_1 指标、编码比率 β 和数据占用空间 W ;

(5) 判断是否有 $M=N$, 若不成立, 令 $M=M+1$, 返回步骤(2); 若成立, 流程结束。

1.3 模型改进

用户的日负荷曲线中含有一些关键用电特征, 能够有效反映出用户的用电特点, 如日最大负荷、日峰谷差、日负荷率等^[18]。通过欠完备自编码器对电表数据进行特征提取, 有效减小了存储容量, 提高了计算效率, 但牺牲了部分检测的正确率。为此, 文中结合用户的典型用电特征对模型进行优化, 与编码数据共同作为 BP 神经网络的输入, 以提

升分类的准确率。

特征优选策略 (feature selection strategy, FSS) 是研究用电特性指标与用户用电行为密切程度的重要方法^[13]。该方法基于互信息与相关系数对用电行为特征进行定量分析, 综合考虑用电信息特征对分析性能的有效性和互补性, 构建特征优选评价指标为:

$$J(x) = I'(x; c) \prod_y (1 - |\rho_{xy}|) \quad (6)$$

$$J(Y) = \sum_y J(y) \quad (7)$$

式中: $J(x)$ 为针对单个用电特征 x 的评价值; $I'(x; c)$ 为用电特征 x 与用户类别 c 的归一化互信息; ρ_{xy} 为用电特征 x 和 y 的相关系数; $J(Y)$ 为优选特征子集 Y 的评价函数; $J(y)$ 为优选特征子集中特征 y 的评价函数。

通过构建不同的特征集合, 采用启发式搜索方法, 根据特征优选评价指标可以搜索出最佳的用户用电特征集合。

文中选择的用户用电特征集合为 {日最大负荷, 日最小负荷, 日平均负荷, 日峰谷差, 日峰谷差率, 日负荷率, 峰时耗电率, 谷电系数, 平段用电百分比}, 采用特征优选策略, 从特征集合中筛选出 3 个典型特征, 与编码数据共同作为 BP 神经网络的输入进行网络训练, 流程如图 4 所示。

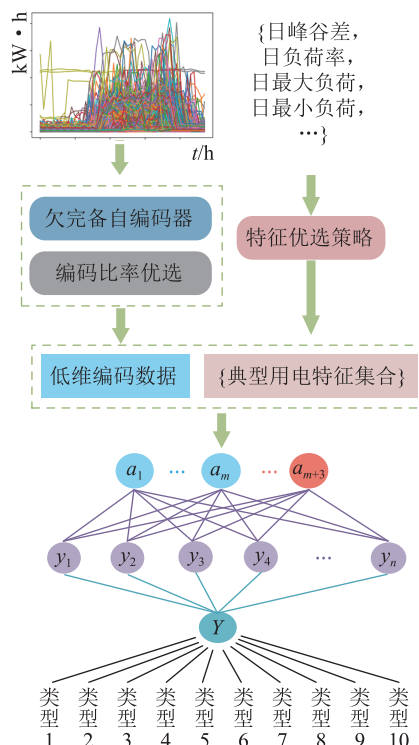


图 4 改进后的用户用电行为分类分析流程

Fig.4 Optimized process of residential electricity consumption behavior classification analysis

2 参数优选实验

本节选择爱尔兰可持续能源局(SEAI)发布的爱尔兰智能电表数据集作为实验对象。该数据集覆盖了4 000多个居民用户的536组日负荷曲线,每条日负荷曲线为48个点,即电表采集频率为30 min/次。采用文献[8]中的居民负荷二次聚类结果作为用户用电分类标签,对欠完备自编码器的编码比率和用户典型用电特征进行优选。文中使用Python3.7.1编写程序,在CPU型号为4200H的Windows10环境下进行了实验。本节还将探讨BP神经网络中激活函数和隐含层节点数对用户用电行为分类准确性的影响,并选择出合适的网络参数以便进行实验对比分析。

2.1 编码比率和用电特征优选

根据文献[8]将爱尔兰居民用户负荷分为10类,用于BP神经网络训练,分类详情见表1。按照图3所示流程对编码比率 β 进行优选,绘制 β - F_1 关系如图5所示。

表1 爱尔兰居民用户负荷类型
Table 1 Types of resident user load in Ireland

| 类别 | 用户负荷特征 |
|----|------------------|
| 1 | 平缓、较低负荷 |
| 2 | 平缓、中等负荷 |
| 3 | 10:00 峰时负荷 |
| 4 | 13:00 峰时负荷 |
| 5 | 16:00 峰时负荷 |
| 6 | 18:00 峰时负荷 |
| 7 | 10:00~16:00 高峰负荷 |
| 8 | 18:00~22:00 高峰负荷 |
| 9 | 10:00~24:00 高峰负荷 |
| 10 | 23:00~7:00 高峰负荷 |

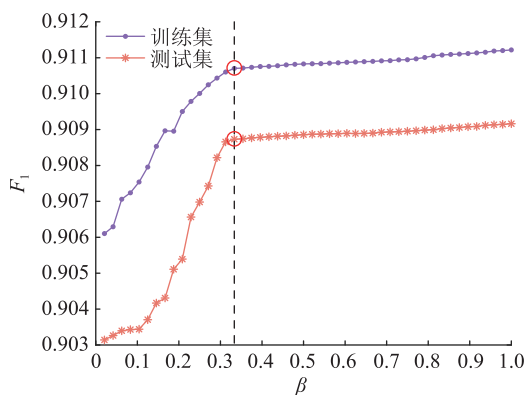


图5 β - F_1 关系

Fig.5 The relationship between β and F_1

由图5可知,训练集和测试集的 F_1 指标均随编

码比率 β 的增大而增大。曲线拐点的位置约为 $\beta = 0.33$,在拐点之前,曲线较为陡峭, F_1 值的增长较快;到达拐点后,曲线变得非常平缓, F_1 指标缓慢增长。因此,文中选择0.33作为最优编码比率,使用欠完备自编码器对原始电表数据编码造成 F_1 值的损失将通过模型优化进行补偿。

在最佳编码比率下应用特征优选策略^[13],得到用户典型用电特征为{日平均负荷,谷电系数,平段用电百分比},与编码数据共同作为BP神经网络的输入进行网络训练。

2.2 BP神经网络激活函数对结果的影响分析

本节通过对比不同隐含层节点数下,不同激活函数对用户用电行为分类准确率的影响。选取隐含层节点数为10—80,选用的激活函数分别为S型函数(sigmoid)、线性整流函数(relu)、双曲正切函数(tanh)和近似平滑函数(softplus)。实验结果如图6所示。

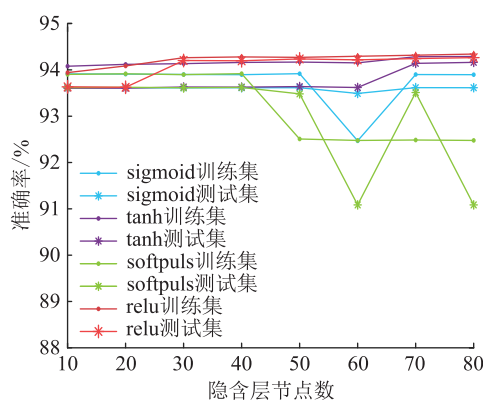


图6 不同激活函数下的准确率对比

Fig.6 Accuracy comparison under different kernel functions

由图6可以看出,采用不同的激活函数,用户用电行为分类的准确率不同。softplus函数和sigmoid函数效果略差,relu函数和tanh函数表现较好,在训练集上的准确率最高分别为94.34%和94.29%。然而,tanh函数在测试集中表现略逊于relu函数。整体来看,relu函数的准确率最高,较其他激活函数更适合进行用户用电行为分类分析。

2.3 BP神经网络隐含层节点数对结果的影响分析

由上述实验可知,采用relu作为网络的激活函数准确率较高,因此本节选用relu作为网络的激活函数。在此基础上,通过改变网络的隐含层节点数目,探究隐含层节点数对用户用电行为分类准确性的影响,并选择最优的隐含层节点数。实验结果如图7所示。

由图7可以看出,设置不同的隐含层节点数目,

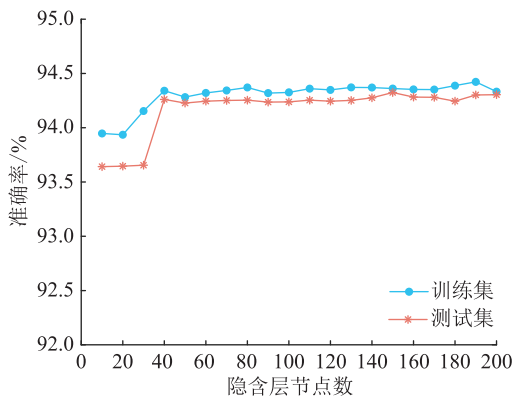


图7 不同隐含层节点数下的准确率对比
Fig.7 Comparison of accuracy with different hidden layer node numbers

用户用电行为分类的准确率不同。在隐含层节点数达到40之前,训练集和测试集的准确率随着隐含层节点数的增加而增大;当隐含层节点数达到40之后,训练集准确率的增长变得非常缓慢,而测试集准确率平稳波动、基本不变,此时增加隐含层节点对预测的准确率影响不大,反而会导致网络更加复杂,增加训练时间和测试时间。因此,隐含层节点数取40较为合适。

3 实验结果与性能分析

根据参数优选的结果,设置欠完备自编码器的编码比率为0.33,用户典型用电特征集为{日平均负荷,谷电系数,平段用电百分比},BP神经网络的激活函数为relu函数,隐含层节点数为40,用户用电行为分类结果如图8所示。文中还与BP神经网络、随机森林(random forest, RF)、支持向量机(support vector machine, SVM),以及BP神经网络与欠完备自编码器的组合方法进行对比,实验结果见表2,效果对比见表3。

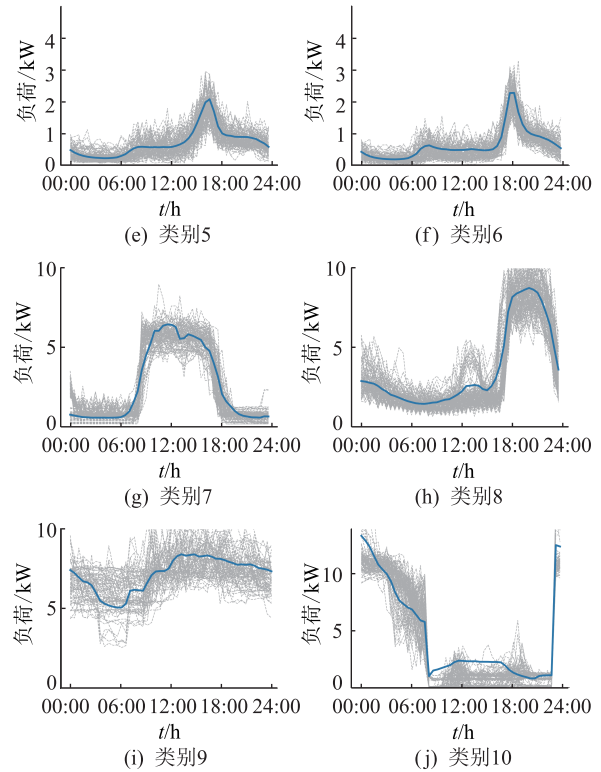
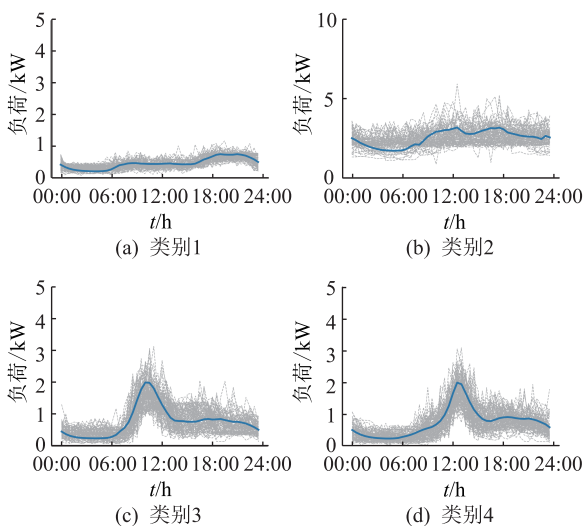


图8 用户用电行为分类结果
Fig.8 Classification results of residential electricity consumption behavior

表2 实验结果对比

| 方法 | 训练集准确率/% | 测试集准确率/% | 训练时间/s | 测试时间/s | 数据占用空间/MB |
|--------|----------|----------|--------|--------|-----------|
| BP | 90.53 | 90.01 | 20.14 | 0.22 | 14 521 |
| BP+UAE | 89.54 | 89.12 | 16.25 | 0.14 | 4 213 |
| RF | 92.01 | 91.85 | 21.56 | 0.24 | 14 521 |
| SVM | 91.21 | 90.89 | 18.86 | 0.19 | 14 521 |
| 文中方法 | 94.34 | 94.26 | 17.64 | 0.16 | 5 120 |

表3 方法效果对比

| 方法 | 训练集准确率 | 测试集准确率 | 训练时间 | 测试时间 | 数据占用空间 |
|--------|--------|--------|---------|---------|---------|
| BP | 4.21 ↑ | 4.72 ↑ | 12.41 ↓ | 27.27 ↓ | 64.74 ↓ |
| BP+UAE | 5.36 ↑ | 5.77 ↑ | 8.55 ↑ | 14.29 ↑ | 21.53 ↑ |
| RF | 2.53 ↑ | 2.62 ↑ | 18.18 ↓ | 33.33 ↓ | 64.74 ↓ |
| SVM | 3.43 ↑ | 3.71 ↑ | 6.47 ↓ | 15.79 ↓ | 64.74 ↓ |

注:表内结果为文中方法效果提升(或降低)的百分比。

由表2和表3可知,使用原始电表数据结合BP神经网络方法进行用户用电行为分类分析的各项指标均表现较差,预测准确率和效率较低,数据占用空间较大。当对电表数据进行欠完备自编码器编码后,程序的训练时间、测试时间以及数据占用空间明显下降,但会损失部分预测准确率,在训

训练集和测试集上损失的准确率分别为 0.99% 和 0.89%。RF 方法使分类的准确性有所提升,在训练集和测试集上分别为 92.01% 和 91.85%,但所需的训练时间和测试时间更久。SVM 方法虽然训练时间和测试时间较短,但准确性没有明显提高。文中方法通过欠完备自编码器与用户典型特征相结合,在训练集与测试集上的准确率分别达到 94.34% 和 94.26%,比传统 BP 神经网络分别提高了 4.21% 和 4.72%,在 5 种方法中表现最好。同时,数据占用空间降低了 64.74%,训练时间和预测时间也显著下降,分别为 17.64 s 和 0.16 s。结果表明,文中所提方法在节约数据存储空间的同时,能够有效提升用户用电行为分类的准确率和计算效率。

4 结语

文中从用户用电行为分类分析在智能电表的应用角度出发,兼顾分类的准确率与数据存储空间,提出了基于欠完备自编码器的用电行为分类分析方法。在爱尔兰数据集上的实验结果表明,所提方法能在节约数据存储空间的同时,有效提升用户用电行为分类的准确率,证明了该方法的正确性与有效性。

目前,人工智能技术与智能电网的联系日益紧密,深度学习算法为传统电网中的难题提供了新的解决方案。文中在用户用电行为分类中,使用了传统的 BP 神经网络作为分类算法,下一阶段将对深度学习算法进行研究,进一步提升模型的预测效果。

本文得到国网江苏省电力有限公司科技项目“面向需求响应的居民分路负荷参与电网互动关键技术研究及装置研发”(J2018039)资助,谨此致谢!

参考文献:

- [1] 夏水斌,张芹,谢玮,等. 电力用户用电信息采集系统建设研究[J]. 自动化与仪器仪表,2018(10):48-50.
XIA Shuibin,ZHANG Qin,XIE Wei,et al. Research on the construction of power user information acquisition system[J]. Automation & Instrumentation,2018(10):48-50.
- [2] 王思彤,周晖,袁瑞铭,等. 智能电表的概念及应用[J]. 电网技术,2010,34(4):17-23.
WANG Sitong,ZHOU Hui,YUAN Ruiming,et al. Concept and application of smart meters[J]. Power System Technology,2010,34(4):17-23.
- [3] 余鹤,夏水斌,董重重,等. 采集终端与智能电表批量采集任务处理技术研究[J]. 仪表技术,2019(5):25-28.
YU He,XIA Shuibin,DONG Chongchong,et al. Study on batch acquisition task processing technology of acquisition terminal and intelligent electricity meter[J]. Instrumentation Technology,2019(5):25-28.

- [4] 王继业,季知祥,史梦洁,等. 智能配用电大数据需求分析与应用研究[J]. 中国电机工程学报,2015,35(8):1829-1836.
WANG Jiye,JI Zhixiang,SHI Mengjie,et al. Demand analysis and application research of intelligent TV big data[J]. Proceedings of the CSEE,2015,35(8):1829-1836.
- [5] 丁晓,孙虹,郑海雁,等. 基于配用电大数据的短期负荷预测[J]. 电力工程技术,2018,37(3):21-27.
DING Xiao,SUN Hong,ZHENG Haiyan,et al. Short-term load prediction based on distributed big data[J]. Electric Power Engineering Technology,2018,37(3):21-27.
- [6] 陈焯,吴浩,史俊祎,等. 奇异值分解方法在日负荷曲线降维聚类分析中的应用[J]. 电力系统自动化,2018,42(3):105-111.
CHEN Ye,WU Hao,SHI Junyi,et al. Application of singular value decomposition method in daily load curve dimensionality reduction clustering analysis[J]. Automation of Electric Power Systems,2018,42(3):105-111.
- [7] 王帅,杜欣慧,姚宏民,等. 面向含多种用户类型的负荷曲线聚类研究[J]. 电网技术,2018,42(10):3401-3412.
WANG Shuai,DU Xinhui,YAO Hongmin,et al. Research on load curve clustering with multiple user types[J]. Power System Technology,2018,42(10):3401-3412.
- [8] 朱文俊,王毅,罗敏,等. 面向海量用户用电特性感知的分布式聚类算法[J]. 电力系统自动化,2016,40(12):21-27.
ZHU Wenjun,WANG Yi,LUO Min,et al. Distributed clustering algorithm for perception of power consumption characteristics of mass users[J]. Automation of Electric Power Systems,2016,40(12):21-27.
- [9] 吴郅君,殷新博,陈中,等. 基于模糊聚类曲线相似度的负荷用户识别方法[J]. 电力工程技术,2019,38(3):151-156.
WU Zhijun,YIN Xinbo,CHEN Zhong,et al. Load user identification method based on fuzzy clustering curve similarity[J]. Electric Power Engineering Technology,2019,38(3):151-156.
- [10] 冯志颖,唐文虎,吴青华,等. 考虑负荷纵向随机性的用户用电行为聚类方法[J]. 电力自动化设备,2018,38(9):39-44,53.
FENG Zhiying,TANG Wenhua,WU Qinghua,et al. Clustering method of user electricity behavior considering load longitudinal randomness[J]. Electric Power Automation Equipment,2018,38(9):39-44,53.
- [11] 林顺富,田二伟,符杨,等. 基于信息熵分段聚合近似和谱聚类的负荷分类方法[J]. 中国电机工程学报,2017,37(8):2242-2253.
LIN Shunfu,TIAN Erwei,FU Yang,et al. Load classification method based on information entropy piecewise polymerization approximation and spectral clustering[J]. Proceedings of the CSEE,2017,37(8):2242-2253.
- [12] 栾开宁,鲍敏,易永仙,等. 基于用电模式数的大用户负荷短期预测技术研究[J]. 电力工程技术,2018,37(3):33-37.
LUAN Kaining,BAO Min,YI Yongxian,et al. Study on short-term prediction technology of large user load based on power consumption mode number[J]. Electric Power Engineering

- Technology, 2018, 37(3):33-37.
- [13] 陆俊,朱炎平,彭文昊,等. 智能用电用户行为分析特征优选策略[J]. 电力系统自动化, 2017, 41(5):58-63, 83.
LU Jun, ZHU Yanping, PENG Wenhao, et al. Optimization strategy of intelligent user behavior analysis characteristics[J]. Automation of Electric Power Systems, 2017, 41(5):58-63, 83.
- [14] 陆俊,陈志敏,龚钢军,等. 基于极限学习机的居民用电行为分类分析方法[J]. 电力系统自动化, 2019, 43(2):97-104.
LU Jun, CHEN Zhimin, GONG Gangjun, et al. Classification and analysis method of residential electricity use behavior based on extreme learning machine[J]. Automation of Electric Power Systems, 2019, 43(2):97-104.
- [15] 史常凯,张波,盛万兴,等. 灵活互动智能用电的技术架构探讨[J]. 电网技术, 2013, 37(10):2868-2874.
SHI Changkai, ZHANG Bo, SHENG Wanxing, et al. Discussion on the technical framework of flexible interactive intelligent electricity use[J]. Power System Technology, 2013, 37(10):2868-2874.
- [16] 王雅思,姚鸿勋,孙晓帅,等. 深度学习中的自编码器的表达能力研究[J]. 计算机科学, 2015, 42(9):56-60.
WANG Yasi, YAO Hongxun, SUN Xiaoshuai, et al. Research on the expression ability of self-encoder in deep learning[J]. Computer Science, 2015, 42(9):56-60.
- [17] HERVÉ A, WILLIAMS L J. Principal component analysis[J]. Wiley Interdisciplinary Reviews Computational Statistics, 2010, 2(4):433-459.
- [18] 宗柳,李扬,王蓓蓓. 计及需求响应的多维度用电特征精细挖掘[J]. 电力系统自动化, 2012, 36(20):54-58.
ZONG Liu, LI Yang, WANG Beibei. Detailed mining of multi-dimensional electricity characteristics based on demand response[J]. Automation of Electric Power Systems, 2012, 36(20):54-58.

作者简介:



黄奇峰

黄奇峰(1968),男,学士,研究员级高级工程师,从事电力负荷柔性调控与电测量技术等方向研究工作(E-mail:hqfyqhqy@126.com);

杨世海(1976),男,硕士,研究员级高级工程师,从事需求响应、智能用电与电测量等方向研究工作;

邓欣宇(1996),男,硕士在读,研究方向为人工智能在配用电系统的应用。

Classification analysis method for electricity consumption behavior based on undercomplete autoencoder

HUANG Qifeng¹, YANG Shihai¹, DENG Xinyu², CHEN Haiwen², WANG Shouxiang²

(1. State Grid Jiangsu Electric Power Co., Ltd. Research Institute, Nanjing 211103, China;

2. Key Laboratory of Smart Grid of Ministry of Education, Tianjin University, Tianjin 300072, China)

Abstract: In view of the complex and changeable power consumption behavior of users under the background of power big data and the difficulty in analysis, a classification and analysis method of power consumption behavior of users based on undercomplete auto-encoder is proposed. Firstly, the data of intelligent electricity meters are encoded by an undercomplete auto-encoder to extract the features of the original data, and the back-propagation (BP) neural network is used to classify and analyze the user's electricity consumption behavior. Then, the optimal coding ratio is selected, and the typical user electricity characteristics are taken as the input of the neural network to improve the classification accuracy. Finally, a simulation experiment is carried out on smart meters in Ireland data sets, compared with directly using the BP neural network analysis, the proposed method not only can improve the accuracy of detection, help electric power company to better grasp the power law of auxiliary demand response, but also can significantly reduce the running time of the algorithm.

Keywords: undercomplete autoencoder; electricity consumption behavior analysis; demand response; feature mining; intelligent electricity consumption; smart meter

(编辑 陈静)