

doi: 10.7690/bgzd.2025.03.006

一种电力用户用电特征数据挖掘方法

李 雄, 吴方权, 汤成佳

(贵州电网有限责任公司信息中心, 贵阳 550003)

摘要: 针对目前电力需求响应分析中存在缺乏对不同环境和激励信号下用户响应行为的分析和预测的问题, 提出一种基于电力用户用电行为数据挖掘方法。通过构建并分析基于激励的需求响应体系结构, 且基于现有的用户响应成本抽象公式, 建立用户响应灵活度模型; 同时提出一种双层长短时记忆网络识别用户响应行为模型, 将所提模型与随机森林(random forest, RF)、支持向量机(support vector machines, SVM)、递归神经网络(recurrent neural network, RNN)、长短时记忆(long short-term memory, LSTM)等模型进行对比。结果表明: 所提模型性能优异, 准确率为 94.83%, F1 分数为 95.45%, 品质因数为 39.42%, 可对电力安全运行管理的发展提供一定借鉴。

关键词: 电力系统; 需求响应; 深度学习; 长短时记忆; 行为识别

中图分类号: MT7 **文献标志码:** A

Power Consumption Characteristic Data Mining Method for Power User

Li Xiong, Wu Fangquan, Tang Chengjia

(Information Center of Guizhou Power Grid Company Limited, Guiyang 550003, China)

Abstract: In order to solve the problem of lack of analysis and prediction of user's response behavior under different environments and incentive signals in the current power demand response analysis, a data mining method based on power user's power consumption behavior is proposed. Construct and analyze the incentive-based demand response architecture, establish a user response flexibility model based on the existing abstract formula of user response cost, and propose a double-layer longshort-term memory network to identify user response behavior model; The proposed model is compared with random forest (RF), support vector machines (SVM), recurrent neural network (RNN) and long short-term memory (LSTM). The results show that the proposed model has excellent performance, the accuracy rate is 94.83%, the F1 score is 95.45%, and the quality factor is 39.42%, which can provide a reference for the development of safe operation and management of electric power.

Keywords: power system; demand response; deep learning; long short-term memory; behavior recognition

0 引言

随着智能电表^[1-2]的大规模普及和应用以及电网的精益管理带来的配电数据的爆炸式增长, 电力用户的准确分类和特征分析已成为电网公司的一项重要任务。此外, 基于网络、大数据、物联网、通信^[3-5]等技术, 可根据用户需求响应的特点, 充分挖掘用户需求响应潜力, 提高需求响应的实施效率, 从而为制定需求侧的管理方案和优化电力价格包提供依据。

随着经济和技术的发展, 电力用户的类型和行为复杂多样。电力用户数据显示出高维和海量数据的特点, 这使得对用户进行分类和分析响应特征变得更加困难^[6]; 因此, 高效处理复杂用户数据和准确分析功耗特性非常重要。文献[7]提出了基于用户电力特征的分割网络模型, 可基于卷积神经网络对电力用户用电特征进行编码, 从而分析用户在相邻时段内用电记录的相关性。文献[8]研究了电力用户

对供电可靠性敏感度的评价与分类方法, 提出基于层次分析算法和模糊 C 均值聚类算法的电力用户敏感度评价算法以及基于两者的综合评价方法。文献[9]提出一种基于自动编码器与改进模糊 C 均值聚类算法结合技术的用户响应特征分析方法。目前大部分研究在分析用户在需求响应中的行为时大多侧重于对用户负载的预测, 缺乏对不同环境和激励信号下用户响应行为的分析和预测。

为改善上述问题, 笔者研究基于激励的需求响应体系结构, 并基于现有的用户响应成本抽象公式, 分析了用户的响应灵活度, 为用户的响应行为识别提供支持。同时, 提出了一种长短时记忆(LSTM)网络识别用户响应行为模型。

1 基于激励的需求响应分析模型

1.1 定性分析

基于激励的需求响应分析模型实施过程如图 1

收稿日期: 2024-07-11; 修回日期: 2024-08-14

第一作者: 李 雄(1994—), 男, 贵州人, 硕士。

所示。负荷服务商可以整合需求侧的资源，作为独立实体参与电力市场。在实施需求响应时，根据用户的状态向其发送不同的激励措施，从而实现目标响应过程。

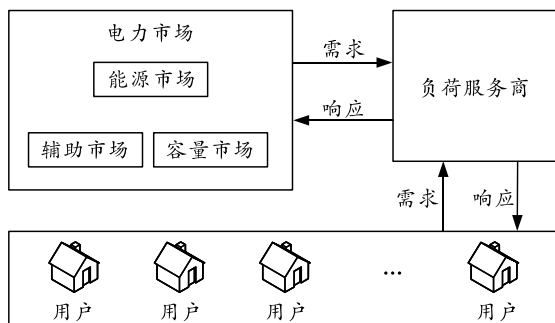


图 1 基于激励的需求响应分析模型实施过程

在能源市场中，负荷服务商需要参与实时电力交易或购买辅助服务，以弥补日前市场获得的电力与实际电力之间的偏差。此外，在辅助市场和容量市场，负荷服务商可以通过需求响应提供负载均衡。一般情况下，在可用期内未调度的所有负载管理，包括 4 种需求响应类型：有限灾难恢复、延长夏季灾难恢复、年度灾难恢复和容量性能灾难恢复所涉及的负载必须进行强制性测试，以证明其能够满足容量要求。如果负荷服务商提供的负载能力太小，则利润将受损；相反，负荷服务商可能会因未能实现目标而受到相应惩罚。当负荷服务商整合需求侧资源参与电力市场的常规业务时，需要对用户的响应行为进行精确预测，从而根据每个需求响应服务中的响应要求向目标用户提供不同的需求激励。基于激励的需求响应可以使负荷服务商的调整更加灵活，有助于提高其业务发展的准确性。对于每个用户，从负荷服务商收到的激励值可以相同或不同，这样可使用户更多地参与需求响应业务，进一步提高电力服务整体效率。

需注意，在基于激励的需求响应的整个业务流程中，用户对不同激励的响应受到许多因素的影响；为此，可以使用响应灵活度来描述不同用户的响应行为。对于不同的用户，响应灵活度可能不同；对于相同的用户，在不同的外部环境下，响应灵活度也可能发生变化。通常，用户的响应灵活度主要受可替代物、负荷状态、用电支出比例以及外部环境等因素影响。首先，可替代物指当电费高昂时，用户将以其他设备取代电力设备。如燃气灶将取代电磁炉，燃气热水器将取代电热水器，从而降低用户整体成本。在这种情况下，用户的响应灵活度将很

高；其次，如果用户当前负载较高且可中断或可转移负载相对较大，则用户将更有可能响应激励，并且响应灵活度将较高；接着，当用户的经济状况较好时，电费在其总支出中所占的比例较小，用户可能对激励不太敏感，相反，用户可能对激励更敏感；最后，外部环境也将影响用户响应。例如，在夏季中午的高温下，用户的响应灵活度可能较低；因此，如果负荷服务商希望鼓励用户更换或关闭空调以通过激励措施降低负荷，则需要较高的成本。与之相反，晚上随着外部环境温度的降低，用户的响应灵活度可能会增加，此时用户可能更愿意以更低的价格响应负载降低需求。

1.2 定量分析

对基于激励的需求响应分析模型进行定量分析。首先，用户响应灵活度定义如下：

$$E_R = \frac{\Delta R / R}{\Delta I / I} = \frac{\Delta R}{\Delta I} \cdot \frac{I}{R} \quad (1)$$

式中： E_R 为响应灵活度； ΔR 为负荷服务商需求响应偏差； ΔI 为负荷服务商激励偏差； R 为用户在需求响应中的响应量； I 为用户收到的激励量。

考虑到与需求响应将伴随一定程度的舒适度损失。为此，笔者将与需求响应的成本函数定义为：

$$U(R) = \beta R^2 / 2 + \alpha R \quad (2)$$

式中： $U(R)$ 为用户参与需求响应的成本函数； α 和 β 为成本参数，可以反映具有不同参数的不同用户的响应行为。一般来说，当用户收到的激励可以补偿舒适度的损失时，用户更愿意做出响应。假定负荷服务商希望用户的响应为 R ，发送给用户的激励 I 应该大于或等于用户的响应成本。本文中，假设当用户收到的激励量高于或等于响应成本时，用户愿意响应。同时，并非所有用户都希望在收到的激励等于舒适度损失时做出反应（当收到的激励高于舒适度损失的时候，用户可能会做出反应），但笔者可以通过调整参数 α 和 β 来找到用户愿意执行响应的临界点。应注意，用户的实际最大响应不应超过当前负荷；因此，有必要对用户的响应量添加约束。在这种情况下，用户的响应和激励之间的关系可以表示为如下目标函数：

$$\begin{aligned} U_{ad}(R) &= \beta_{ad}(R + \varepsilon)^2 / 2 + \alpha_{ad}(R + \varepsilon) \\ \text{s.t. } R &\leq L \end{aligned} \quad (3)$$

式中： $U_{ad}(R)$ 为用户的响应成本目标函数； L 为用户的当前负荷； ε 为用户响应的噪声； α_{ad} 和 β_{ad} 为响

应参数，其值取决于用户的响应行为。同时，由于用户响应的噪声是低阶和随机的，笔者在分析用户的需求灵敏度时忽略了噪声。最终，用户的响应灵敏度可更新如下：

$$E_R = \frac{dR}{dU_{ad}} \cdot \frac{U_{ad}}{R} = \frac{\beta_{ad}R + 2\alpha_{ad}}{2\beta_{ad}R + 2\alpha_{ad}} \quad (4)$$

式中： E_R 为用户的响应灵敏度，反映了用户对激励变化的反应的敏感性。

2 基于 LSTM 的用户响应行为预测模型

考虑到用户激励的需求响应行为受多种因素影响，传统的线性或非线性近似模型都很难用时间序列描述隐含关系。LSTM 网络具有强大的时间序列相关数据处理能力，可通过单元存储历史数据，并通过遗忘门删除较少相关数据，从而准确预测时间序列下的用户响应行为。

考虑到单个 LSTM 单元只能捕获用户响应行为数据长期上下文关系，笔者使用双层 LSTM 特征提取模型学习短期用户响应行为数据关系，从而有效提高模型的分类和回归能力。双层 LSTM 网络结构如图 2 所示，其包含 2 个相反方向的 LSTM 隐藏层：前向和后向 LSTM 单元。令时间序列数据为 $x=(x_1, x_2, \dots, x_n)$ ，则双层 LSTM 从开始到结束同时遍历特征向量，前向神经元隐藏层输出序列 $\vec{h}_t$ 和后向神经元隐藏层输出序列 $\bar{h}_t$ 可描述为：

$$\vec{h}_t = (\vec{h}_1, \vec{h}_2, \dots, \vec{h}_n) ; \quad (5)$$

$$\bar{h}_t = (\bar{h}_1, \bar{h}_2, \dots, \bar{h}_n) \quad (6)$$

式中 n 为特征序列中包含特征的个数。

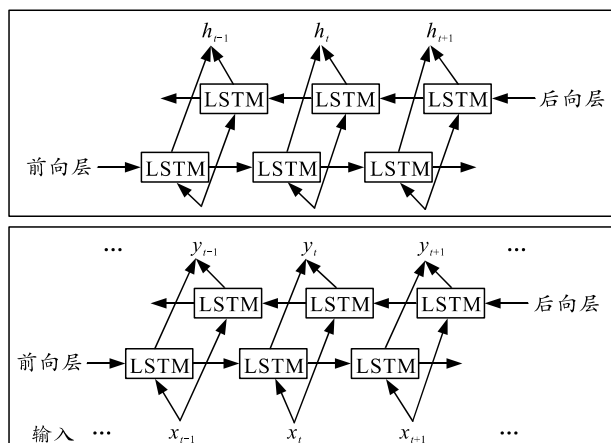


图 2 双层 LSTM 网络结构

进一步，双向 LSTM 网络隐藏层的编码输出 y_t 可以通过级联正向和反向输出求解：

$$y_t = [\vec{h}_t, \bar{h}_t] ; \quad (7)$$

$$\vec{h}_t = \sigma(W_{\vec{h}_t} x_t + W_{\vec{h}_{t-1}} \vec{h}_{t-1} + b_{\vec{h}}) ; \quad (8)$$

$$\bar{h}_t = \sigma(W_{\bar{h}_t} x_t + W_{\bar{h}_{t-1}} \bar{h}_{t-1} + b_{\bar{h}}) ; \quad (9)$$

$$y_t = W_{y_{\vec{h}_t}} \vec{h}_t + W_{y_{\bar{h}_t}} \bar{h}_t + b_y \quad (10)$$

式中 W 和 b 分别为与 LSTM 的 3 个门相对应的权重向量和偏置。

在训练模型时，可以使用梯度下降算法来训练参数。令 θ 为深度学习网络的参数（权重和偏差）， $L(\theta)$ 为整个网络的损失函数。网络优化过程是找到一个参数 θ ，使得 $L(\theta)$ 最小。梯度下降法以迭代方式使参数朝总损失方向减小的方向更新：

$$\theta_{n+1} = \theta_n - \alpha \partial L(\theta_n) / \partial \theta_n \quad (11)$$

式中： α 为学习率； $\partial L(\bullet)$ 为预测值和实际值之间的均方误差； n 为迭代次数。在每次迭代中，随机选择训练样本的损失函数进行优化，从而加快每次迭代的更新速度。通过随机梯度下降法获得的最小损失可能不代表所有数据的最小损失；因此，结合梯度下降和随机梯度下降的优点，每次计算训练数据的损失函数的一小部分，即执行数据批处理操作。批处理的使用使得每次迭代中优化的参数不会太小，并且可以减少迭代次数以达到收敛，并使收敛结果更接近梯度。

3 实验与分析

3.1 数据集

基于中国某电力公司提供的用户历史用电数据对所提模型进行验证。电力数据收集时间为 2014—2016 年，收集频率为每 10 min 采样一次。

考虑到同一用户一天不同时间的响应灵敏度可能不同，用户在不同时间段的 α_{ad} 和 β_{ad} 值可能不同。为反映用户响应行为中的随机噪声 ε ，首先假定随机噪声服从高斯分布，接着对用户历史用电数据进行聚类分析，并生成 α_{ad} 和 β_{ad} 值。最终，一天内用户响应参数设置如表 1 所示。

表 1 一天内用户响应参数设置

参数	时间段			
	0~6	7~12	13~18	19~24
α_{ad}	1.500	3.000	1.000	1.700
β_{ad}	6.000	5.000	6.000	4.200
标准差	0.212	0.218	0.223	0.198

3.2 实验环境与对比指标

在构建深度预测模型时，为防止模型过拟合，双层 LSTM 模型每层添加概率为 0.2 的 dropout 层。仿真软件环境为 pycharm 搭建算法框架，并由

python 基于 tensorflow 搭建 LSTM 基础网络。同时, 算法运行硬件环境为酷睿 i7 CPU, 内存为 128 G ARM 的联想服务器, 操作系统为 Ubuntu 18.04 64 位, 显卡为 NVIDIA RTX2080Ti 11 G。深度网络学习训练时参数设置如下: 初始学习率设置为 10^{-4} , 数据批量处理大小设置为 16、学习率衰减倍数设置为 0.05、最大迭代次数为 15 000 次。实验时, 测试数据集共包含 20 组需求响应事件, 且各组需求响应事件收到的激励如图 3 所示。

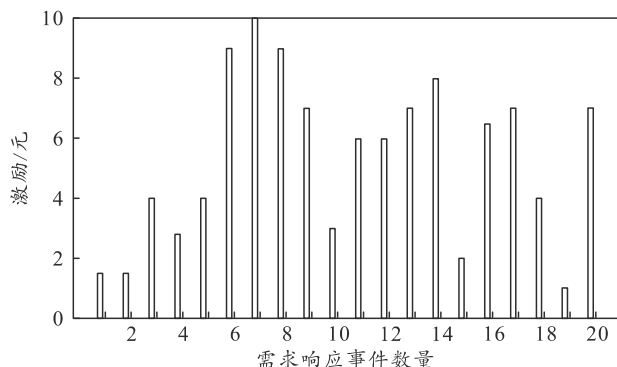


图 3 各组需求响应事件收到的激励

实验分别对比了 RF、SVM、RNN、LSTM 等模型的性能, 从而验证所提模型的优势。对比指标分别选取准确率、F1 分数和品质因数 (figure of merit, FoM)。其中, 准确率表征模型预测准确的程度。F1 得分用来平衡精确率和召回率。FoM 侧重于描述观测和预测值之间的变化, 其定义如下:

$$\text{FoM} = B / (A + B + C + D) \quad (12)$$

式中: A 为将观测到的变化预测为持续性的误差单元数; B 为将观测变化预测为变化的正确单元数; C 为将观测变化预测为错误类别的误差单元数; D 为将观测持续性预测为变化的误差单元数。需注意, 更高的 FoM 值表明更高的一致性。

3.3 对比与分析

表 2 为不同模型在测试数据集上的平均预测结果。可以看出: 与随机森林 (RF) 和支持向量机 (SVM) 等传统机器学习模型相比, 深度学习模型递归神经网络 (RNN) 和 LSTM 准确率提升明显。实验结果表明了深度学习模型可有效学习具有时间依赖的电力用户用电数据。此外, 所提模型表现出了优异性能,

准确率为 94.83%, F1 分数为 95.45%, FoM 为 39.42%。结果表明: 所提模型对电力用户响应行为挖掘具有优异性能, 并验证了 CNN 特征提取和 RF 数据平衡的重要作用。

表 2 不同模型在测试数据集上的平均预测结果

模型	准确率	F1 分数	FoM
RF	0.857 2	0.863 8	0.176 3
SVM	0.868 9	0.879 2	0.174 3
RNN	0.881 4	0.920 5	0.265 2
LSTM	0.893 8	0.924 5	0.286 4
所提模型	0.948 3	0.954 5	0.394 2

4 结论

笔者对电力系统中用户用电行为进行了分析, 提出一种电力用户用电特征数据挖掘方法, 实现基于深度学习探索用户用电激励响应行为。该模型可准确预测用户的响应行为, 为电力运营以及负荷转移的发展提供一定借鉴。

参考文献:

- [1] 熊尉辰, 宋国兵, 李洋, 等. 利用智能电表量测数据的三相四线制配电线路参数辨识[J]. 电力系统自动化, 2022, 46(20): 155-166.
- [2] 张芹, 夏水斌, 许健. 基于智能电表的电能信息采集系统的设计与研究[J]. 自动化仪表, 2022, 43(2): 82-87.
- [3] 毛龙灿, 杨南. 基于大数据背景的皮革人才培养优化研究[J]. 中国皮革, 2021, 50(9): 38-41.
- [4] 杨涛. 互联网时代下皮革行业电商物流体系研究分析[J]. 中国皮革, 2021, 50(8): 82-85.
- [5] 钟建棚, 余少锋, 廖崇阳, 等. 基于云计算的电力设备智能监测系统[J]. 云南师范大学学报(自然科学版), 2022, 42(3): 37-41.
- [6] 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.
- [7] 陈溢. 基于数据分析的电力用户用电特征研究[J]. 自动化仪表, 2022, 43(9): 100-105.
- [8] 张宇轩, 苏娟, 贾涛, 等. 电力用户对供电可靠性敏感度的评价与分类方法[J]. 江苏大学学报(自然科学版), 2021, 42(5): 575-581, 608.
- [9] 王剑锋, 倪家明, 王旭东, 等. 基于 AE-MFCM 技术的电力用户响应特性分析方法[J]. 电力系统及其自动化学报, 2021, 33(4): 47-54.