

引用格式: 谢禄江, 段立, 谭刚, 等. 基于双向 LSTM 的用户用电行为识别[J]. 微电子学与计算机, 2021, 38(5): 36-41.  
[XIE L J, DUAN L, TAN G, et al. Electrical behavior recognition based on bidirectional LSTM[J]. Microelectronics & Computer, 2021, 38(5): 36-41.]

# 基于双向 LSTM 的用户用电行为识别

谢禄江<sup>1</sup>, 段立<sup>1</sup>, 谭刚<sup>2</sup>, 蒋荣<sup>1</sup>, 钟淘淘<sup>1</sup>,  
刘美川<sup>1</sup>, 雷洋<sup>1</sup>, 廖军<sup>3</sup>

(1 国网重庆市电力公司 信息通信分公司, 重庆 400010; 2 重庆智网科技有限公司 信息通信分公司, 重庆 400000;  
3 重庆大学 大数据与软件学院, 重庆 401331)

**摘 要:** 为了解决由非技术性损失所造成的用户用电异常问题, 本文提出了一种基于双向长短时记忆神经网络 (Bi-LSTM) 的用户异常用电行为检测方法. 该方法首先采用插值法处理用电缺失数据, 并通过分位数归一化平滑用电异常行为值的不对称分布性. 然后, 结合 LSTM 神经网络单元和双向网络构建 Bi-LSTM 模型, 用于获取用户异常用电行为中的隐含关系. 最后, 采用交叉熵确定最优参数, 从而识别出异常用电用户行为. 实验表明, 该方法的识别性能显著优于其他模型, 并且结合国家电网的实际数据验证了该方法的准确性和稳定性.

**关键词:** 非技术损失; 异常用电行为; Bi-LSTM; 分位数归一化

**中图分类号:** TP181; TM933.4

**文献标识码:** A

**文章编号:** 1000-7180(2021)05-0036-06

## Electrical behavior recognition based on bidirectional LSTM

XIE Lu-jiang<sup>1</sup>, DUAN Li<sup>1</sup>, TAN Gang<sup>2</sup>, JIANG Rong<sup>1</sup>, ZHONG Tao-tao<sup>1</sup>,  
LIU Mei-chuan<sup>1</sup>, LEI Yang<sup>1</sup>, LIAO Jun<sup>3</sup>

(1 State Grid Chongqing Information & Communication Company, Chongqing 400010, China;

2 Chongqing Smart Power Grid Technology Co., Ltd, Chongqing 400000, China;

3 School of Big Data & Software Engineering, Chongqing University, Chongqing 401331, China)

**Abstract:** To solve the problem of abnormal user power consumption caused by non-technical losses, a method for detecting abnormal power consumption behavior based on a bidirectional long-term short-term memory neural network (Bi-LSTM) is proposed. In this method, firstly, the missing data is processed by interpolation method, and the asymmetric distribution of abnormal electricity consumption behavior value is smoothed by quantile normalization. Then, the Bi-LSTM model is constructed by combining the LSTM neural network unit and bidirectional network, which is used to obtain the implicit relationship of abnormal electricity consumption behavior of users. Finally, the cross entropy is used to determine the optimal parameters, thereby detecting the abnormal power consumer. The experiments are conducted based on real energy consumption data, and the results show that the detection performance of this proposed method outperforms other methods in terms of accuracy and stability.

**Key words:** non-technical loss; abnormal power consumption behavior; bidirectional long and short-term memory network; quantile normalization

收稿日期: 2020-07-09; 修回日期: 2020-08-28

基金项目: 国家重点研发计划(2017YFB1401702); 重庆市人工智能技术创新重大主题专项课题(cstc2017rgzn-zdyf0051); 国家自然科学基金面上项目(61977012)

## 1 引言

根据全球银行的统计,2017 年全球超过 88% 的人口都有使用电力的需求. 电力系统是指由发电机、输配电线路、变配电所以及各种用户用电设备连接起来所构成的有机整体. 其中,电力传输和分配中的电量损失是全世界电力公司面临的重要问题. 电量损失通常分为技术损失(Technical Loss, TL)和非技术损失(Non-Technical Loss, NTL). TL 是由电力系统本身的材料(例如电缆,连接器和电源变压器)中功率耗散所引起的,一般很难用信息技术手段避免,而 NTL 是由于电力盗窃、仪表计费或计量错误和非法接线等原因造成的电量损失. 异常用电行为(例如,盗电)主要是由外部因素引起,进而造成 NTL. 因此,异常用电行为是可以通过信息技术手段来识别从而规避的<sup>[1]</sup>.

据调查,这些由 NTL 造成的异常用电行为可能会导致电力公司的收入损失<sup>[2-3]</sup>. 如在美国,公共事业公司因盗窃而损失占其年收入的 0.5% 至 3.5%<sup>[2]</sup>. 在一些发展中国家,因用电异常行为造成的收入损失甚至更大. 近年来,在电力大数据分析背景下,基于数据智能的异常用电行为识别方法已广泛应用于智能电网中,如电量预测和故障检测等<sup>[3]</sup>. 文献[4]采用随机森林(RF)方法对电力负荷进行预测,得到了较好的检测结果. 孙向阳等人<sup>[5]</sup>提出了基于 XGBoost 算法的窃电用户识别模型,对用户特征进行分类,平均准确率达到了 84.41%. 通过以上研究,我们发现,大部分学者已经在用户异常用电检测方法上取得了一定的成果,但他们所使用的大多是传统的机器学习方法,依赖于规则知识. 相对于传统的机器学习方法,深度学习有着强大的自学能力. 目前,深度学习在用电异常行为检测中起着至关重要的作用<sup>[6]</sup>. 文献[7]提出了基于改进的反向传播神经网络(Improved-BP),构建了用户用电异常行为分析模型,实现了异常用电行为的检测. 文献[8]提出了一种基于深度学习的长短时记忆(LSTM)方法,利用 LSTM 神经网络提取用户用电数据特征,并通过全连接网络(FCN)进行多层特征匹配,实现了用户用电模式的异常检测.

鉴于以上研究,为了避免 NTL 所造成的损失,促使用电异常用户诚信用电,本文提出了一种基于双向长短时记忆网络(Bi-LSTM)的用户异常用电行为识别方法,该方法采用分位数归一化处理不对称分布数据,并通过融合 LSTM 神经网络单元和双

向网络来构建 Bi-LSTM 模型,以更好地获取用户用电异常行为中所隐含的特征关系,从而实现对用户用电行为的检测,提高异常用电行为识别的准确率和稳定性. 本文通过真实的用电数据来评估所提出的方法,并与其他相关方法进行比较. 在实际应用中,将识别出的异常用电用户纳入银行金融征信范围,限制其社会经济活动,从而减少社会经济损失,确保电网的安全和稳定.

## 2 用电行为识别模型

本文所提出的方法的主要目的是基于中国国家电网公司发布的真实用电数据,检测出用户用电异常行为,核心思想是将用户用电异常行为识别问题转化为一个有监督的分类问题,然后构建一个能够有效处理用电数据的双向长短时记忆神经网络(Bi-LSTM)模型,以实现用电异常行为的识别. 该方法具体流程如图 1 所示.

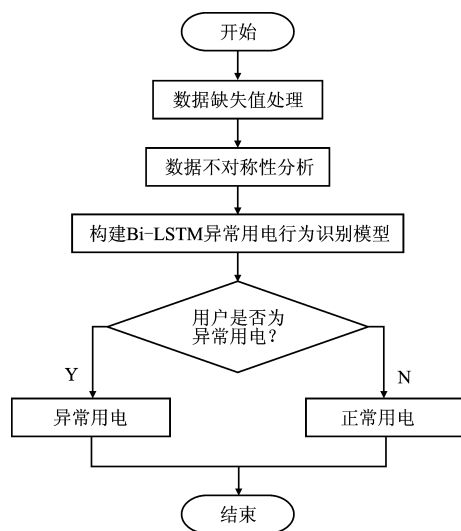


图 1 基于 Bi-LSTM 异常用电行为识别流程图

### 2.1 数据差异性分析

#### 2.1.1 缺失值处理

用电数据通常存在数据缺失现象,这主要是由于各种原因所导致的,例如智能电表损坏或故障,数据传输不可靠,系统维护等原因.

为了处理数据中的缺失值,我们采用插值方法来补全缺失值,公式如下:

$$f(x_i) = \begin{cases} \frac{x_{i-1} + x_{i+1}}{2}, & x_i \in \text{NaN}, x_{i-1}, x_{i+1} \notin \text{NaN} \\ 0, & x_i \in \text{NaN}, x_{i-1} \vee x_{i+1} \in \text{NaN} \\ x_i, & x_i \notin \text{NaN} \end{cases} \quad (1)$$

式中,  $x_i$  代表在一段时间(例如一天)内用电数据的特征值. 如果  $x_i$  为空或非数字字符, 我们将其表示为 NaN(NaN 是一个集合).

### 2.1.2 不对称性分析

电力数据存在一种称为非恒定变异的现象, 该现象会导致数据分布不对称, 即分布右侧的变异性很大, 会产生较长的尾部. 由于这种非恒定的变化, 所引起的不对称性会造成深度学习模型中的虚假交互. 为了处理这种不对称分布, 我们采用了分位数均匀归一化<sup>[9]</sup>, 它是一个非线性函数, 独立应用于各类特征数据, 目的是将所有用电数据置于相同的分布中, 即均匀分布在  $(0, 1)$  之间. 假设  $x_i$  为用电数据的第  $i$  个特征值, 其连续累积分布函数  $F(x_i): R \rightarrow [0, 1]$ , 设概率为  $p(0 < p < 1)$ , 有

$$F(x_i) = \begin{cases} 0, & x \leq x_{\min} \\ \frac{x - x_{\min}}{x_{\max} - x_{\min}}, & x_{\min} \leq x \leq x_{\max} = p \\ 1, & x > x_{\max} \end{cases} \quad (2)$$

式中,  $x_{\min}, x_{\max}$  分别表示电力数据的最小值和最大值. 分位数函数  $G$  可表示为:

$$G(p) = \inf\{x \in R; p \leq F(x_i)\} \quad (3)$$

由于分布函数  $F(x_i)$  具有连续且严格单调的性质,  $F$  的反函数  $F^{-1}$  服从同一分布, 故有  $G = F^{-1}$ . 一般来说, 反函数  $F^{-1}$  与分布函数  $F$  可逆, 由此可知分位数函数  $G$  的返回值  $x'_i$ , 如下:

$$x'_i = G(F(x_i)) \quad (4)$$

我们通过每个用户用电数据特征的累积分布来估计  $F$ , 并使用平均数据特征分位数的累积分布来估计  $G$ . 通过变换, 分位数归一化平滑了数据的不对称性, 有助于提高深度学习模型中异常检测的准确性.

## 2.2 用电异常行为识别

### 2.2.1 网络结构

LSTM<sup>[2,8]</sup> 作为一种循环神经网络(RNN)体系结构, 可以学习长期的时序依赖关系. 它解决了现有 RNN 中的梯度爆炸和梯度指数级衰减问题, 能够有效地刻画时序数据中的内在结构, 并很好地解决了许多时序分类问题.

LSTM 的基本网络结构如图 2 所示. LSTM 中每一个区块主要包含一个或多个循环连接的记忆单元(Cell)及 3 个门结构, 即遗忘门(Forget gate), 输入门(Input gate)和输出门(Output gate). LSTM 具有门控制机制, 该机制允许网络“忘记”存储在记

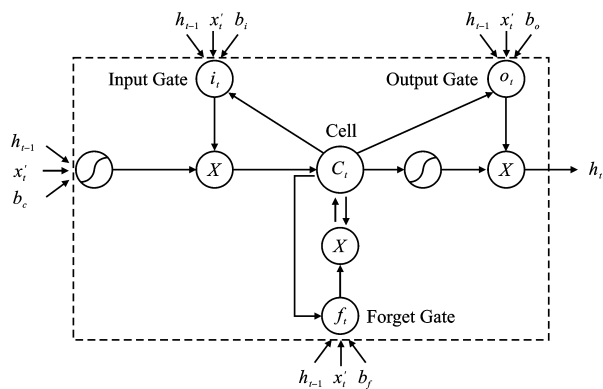


图 2 LSTM 基本网络结构

忆单元中过去输入的状态, 通过门控单元来控制信息的遗忘和更新. 令  $c_t$  表示在  $t$  时刻的记忆单元状态, 通过以下方式更新  $c_t$ :

$$i_t = \sigma(W_{xi}x'_t + W_{hi}h_{t-1} + b_i) \quad (5)$$

$$f_t = \sigma(W_{xf}x'_t + W_{hf}h_{t-1} + b_f) \quad (6)$$

$$o_t = \sigma(W_{xo}x'_t + W_{ho}h_{t-1} + b_o) \quad (7)$$

$$g_t = \tanh(W_{xx}x'_t + W_{hx}h_{t-1} + b_c) \quad (8)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ g_t \quad (9)$$

$$h_t = o_t \circ \tanh(c_t) \quad (10)$$

式中,  $\sigma$  是 Sigmoid 激活函数,  $x \circ y$  表示逐元素乘积.  $x'_t$  表示当前输入向量,  $W_{xi}, W_{hi}, \dots, W_{xx}, W_{hx}$  是线性变换的权重矩阵.  $b_i, b_f, b_o, b_c$  是偏差向量.  $i_t$  是输入门向量,  $f_t$  是遗忘门向量,  $o_t$  是输出门向量,  $g_t$  是状态更新向量,  $h_t$  是输出隐藏状态向量.

### 2.2.2 Bi-LSTM 网络异常用电行为识别模型

对于许多时序分类任务, 利用过去以及未来的上下文信息是有益的. 但是, 传统的 LSTM 网络按时间顺序来处理时序数据, 忽略了过去信息. Bi-LSTM 网络<sup>[10]</sup> 扩展了 LSTM 网络, 引入了第二层结构. 在这一层中, 隐含状态到隐含状态之间以相反的时间顺序进行连接. 因此, Bi-LSTM 模型能够同时利用过去和未来的信息. 在我们的工作中, 建立了基于 Bi-LSTM 用电异常行为的检测模型, 对用户用电是否异常进行分类检测, 如图 3 所示. 模型主要通过级联两层 Bi-LSTM 网络的输出 (即 LSTM 层和反向 LSTM 层), 对隐含状态  $f_t$  添加一个全连接网络 (Fully connected), 并利用 sigmoid 函数提取与用户异常用电行为分类标签相关信息, 即异常用电 (Abnormal) 或正常用电 (Normal).

在以上模型结构中,  $x'_t$  表示当前  $t$  时刻经过平滑处理后的用电数据,  $x'_{t-1}$  和  $x'_{t+1}$  分别表示  $t-1$  和  $t+1$  时刻的用电数据. 对于给定序列中的每个用电

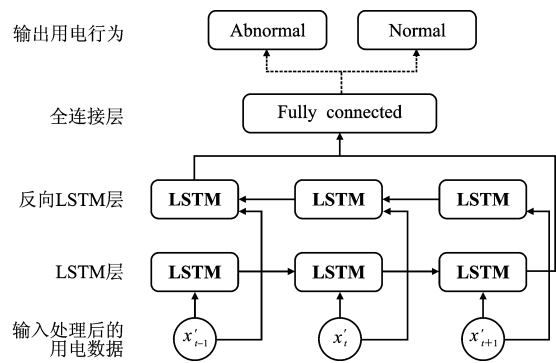


图 3 基于 Bi-LSTM 的用户异常用电行为检测网络结构

信息,由 Bi-LSTM 自动提取时序特征,该模型描述了任意时刻前后所有用电数据的完整信息.

2.3 损失函数

为了提高模型检测性能,本文使用交叉熵损失函数对模型进行度量.通过损失函数  $f_{\text{loss}}$  的计算,对 Bi-LSTM 网络各参数进行调整如下式所示.

$$f_{\text{loss}} = \sum_{i=1}^N y_i \ln \hat{y}_i + (1 - y_i) \ln (1 - \hat{y}_i) \quad (11)$$

式中,  $N$  表示用户的数量,  $y_i$  表示用户  $i$  用电数据真实的分类,  $\hat{y}_i$  表示用户  $i$  用电数据的预测结果(异常为 1,正常为 0).由于充分考虑到一段时间内时序数据点的动态特征信息,Bi-LSTM 模型在训练过程中可以自动学习到复杂动态序列中异常点和正常点的差异,模型经训练后可用于时序异常点的检测,具有较高的性能.因此,训练后的模型能够有效地检测出用电异常行为和正常行为之间的差异,可以准确地检测出用电异常行为,具有较高的准确性.

3 实验分析

3.1 用电数据集

本文采用的数据集为国家电网有限公司真实的用户用电数据,用以验证模型的性能.表 1 列出了该数据集的相关信息,包含 1 035 天内(2014 年 1 月 1 日至 2016 年 10 月 31 日)的 42 372 名用户的用电数据,其中用电异常行为记录占 9%.本实验中,我们将 33 898 条作为训练集,8 474 条为测试集.

表 1 用户用电数据集信息

| 用电数据    | 数量                              |
|---------|---------------------------------|
| 用户用电时间段 | 2014 年 1 月 1 日-2016 年 10 月 31 日 |
| 用户用电总数  | 42 372                          |
| 正常用电数   | 38 757                          |
| 异常用电数   | 3 615                           |

值得注意的是,该数据集包含缺失值和不对称性.因此,我们先利用第 2.1 节中描述的数据差异性分析方法来解决此问题.实验环境以 Python3. 6. 5 为运行平台,并使用 Pycharm 为编程环境.

在本文中,用户用电异常行为问题被认为是二分类任务(即异常或正常),因此,我们采用了准确率(Accuracy)、精确率(Precision)、召回率(Recall)、 $F_1$  值( $F_1$ )和 AUC 五种指标<sup>[6,8]</sup>来评价模型性能.前四个指标是分类指标中最常用的,显示有关分类器性能的信息,从中可以提取更有意义的实验结果.

此外,AUC(Area Under the ROC Curve)是衡量二分类问题中模型稳定性的重要指标.本文以真阳率(True Positive Rate, TPR)作为纵轴,假阳率(False Postive Rate, FPR)作为横轴作图,其中 TPR 与召回率相同.通过 FPR 可画出 ROC 曲线,而 AUC 则是 ROC 曲线下的面积.在 ROC 中,曲线越接近(0,1)点,分类效果就越好,即 AUC 的值越大,表明模型就越好.

3.2 实验结果分析

为了验证 Bi-LSTM 模型的准确性和稳定性,将该模型与 RF<sup>[4]</sup>, XGBoost<sup>[5]</sup>, FF-RNN<sup>[6]</sup>, Improved-BP<sup>[7]</sup>和 LSTM<sup>[8]</sup>等五个模型进行对比实验.本文模型由当天的用电数据预测第二天的用电数据,实现单步时序预测,即时间步长为 1.具体实验结果如下.

3.2.1 准确性对比分析

表 2 给出了不同模型在不同识别指标中的实验对比结果.从表 2 可以看出,Bi-LSTM 模型在各项指标中均优于其他模型,其各指标性能比其他模型都约高 2%~9%.特别地,Bi-LSTM 模型的准确率和召回率均达到 93%,精确率和  $F_1$  值均达到 92%.在用电异常行为识别问题中,Bi-LSTM 模型可以自动学习到用户异常用电的双向隐含关系,从而获得较高的性能,这也意味着该模型能检测到更多真实的用电异常行为.虽然 FF-RNN、RF、XGBoost、Improved-BP 和 LSTM 模型都能识别到当前用户异常用电行为,但它们没有考虑当前用户用电在时序上的上下文信息,从而导致这些模型的检测性能相对于 Bi-LSTM 模型较低.因此,Bi-LSTM 模型可以更好的用于用电异常行为的识别,具有较高的准确性.

表 2 不同模型实验结果对比

| 模型          | 准确率         | 精确率         | 召回率         | $F_1$       |
|-------------|-------------|-------------|-------------|-------------|
| RF          | 0.92        | 0.89        | 0.92        | 0.89        |
| FF-RNN      | 0.92        | 0.85        | 0.92        | 0.88        |
| XGBoost     | 0.92        | 0.90        | 0.92        | 0.90        |
| Improved-BP | 0.904       | 0.91        | 0.90        | 0.91        |
| LSTM        | 0.919       | 0.84        | 0.91        | 0.87        |
| Bi-LSTM     | <b>0.93</b> | <b>0.92</b> | <b>0.93</b> | <b>0.92</b> |

3.2.2 稳定性分析

图 4 描述了不同模型的 ROC 曲线及 AUC 值。由图 4 可知, RF、FF-RNN、XGBoost、Improved-BP、LSTM 在 ROC 曲线下 AUC 的值分别为 81%、50%、81%、75%和 50%,相对于这五种检测模型, Bi-LSTM 模型的 AUC 为 85%,明显高于其他模型。ROC 曲线也更接近于(0,1)这一点,表明 Bi-LSTM 检测模型分类效果优于其他模型,且具有更高的稳定性。

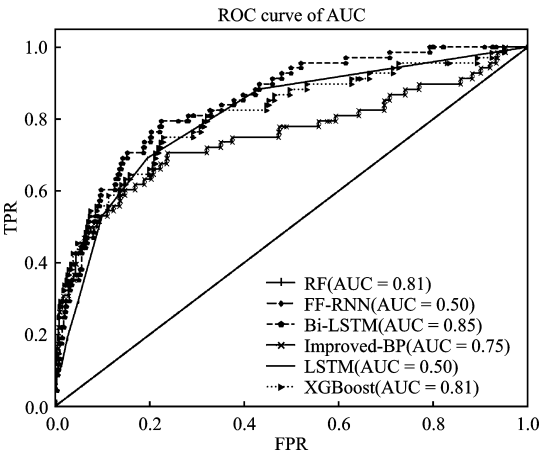


图 4 不同模型的 ROC 曲线及 AUC 值

4 结束语

本文提出了一种基于双向长短时记忆神经网络(Bi-LSTM)模型识别国家电网中用户用电异常行为。通过分位数归一化方法处理不对称用电数据,并结合 LSTM 神经网络单元和双向网络来获取用电行为中的隐含关系,以交叉熵损失函数得到高精度的用电异常行为识别结果。实验结果表明,该方法的检测性能显著优于其他模型,可以有效地满足用户用电异常行为识别的需求,具有较好的准确性和稳定性。在未来的研究中,我们将进一步完善深度学习模型和参数,提高异常用电行为识别性能。

参考文献:

[1] MICHELI G, SODA E, VESPUCCI M T, et al. Big data analytics: an aid to detection of non-technical losses in power utilities[J]. Computational Management Science, 2019, 16 (1-2): 329-343. DOI: 10.1007/s10287-018-0325-x.

[2] GAO Y Q, FOGGO B, YU N P. A physically inspired data-driven model for electricity theft detection with smart meter data[J]. IEEE Transactions on Industrial Informatics, 2019, 15(9): 5076-5088. DOI: 10.1109/TII.2019.2898171.

[3] AHMAD T, CHEN H X, WANG J Y, et al. Review of various modeling techniques for the detection of electricity theft in smart grid environment[J]. Renewable and Sustainable Energy Reviews, 2018, 82: 2916-2933. DOI: 10.1016/j.rser.2017.10.040.

[4] 李婉华, 陈宏, 郭昆, 等. 基于随机森林算法的用电负荷预测研究[J]. 计算机工程与应用, 2016, 52 (23): 236-243. DOI: 10.3778/j.issn.1002-8331.1606-0203.

LI W H, CHEN H, GUO K, et al. Research on electrical load prediction based on random forest algorithm [J]. Computer Engineering and Applications, 2016, 52(23): 236-243. DOI: 10.3778/j.issn.1002-8331.1606-0203.

[5] 孙向阳. 基于 XGBoost 算法的窃电用户行为研究 [D]. 兰州: 兰州交通大学, 2018.

SUN X Y, Research on stealing user behavior based on XGBoost algorithm [D]. Lanzhou: Lanzhou Jiaotong University, 2018.

[6] NABIL M, ISMAIL M, MAHMOUD M, et al. Deep learning-based detection of electricity theft cyber-attacks in smart grid AMI networks[M]//ALAZAB M, TANG M. Deep Learning Applications for Cyber Security. Cham: Springer, 2019.

[7] 田野, 张程, 毛昕儒, 等. 运用 PCA 改进 BP 神经网络的用电异常行为检测[J]. 重庆理工大学学报(自然科学), 2017, 31(8): 125-133. DOI: 10.3969/j.issn.1674-8425(z).2017.08.021.

TIAN Y, ZHANG C, MAO X R, et al. Research on abnormal behavior of power consumption based on BP neural network with PCA[J]. Journal of Chongqing University of Technology (Natural Science), 2017, 31 (8): 125-133. DOI: 10.3969/j.issn.1674-8425(z).2017.08.021.

[8] 赵文清, 沈哲吉, 李刚. 基于深度学习的用户异常用电模式检测[J]. 电力自动化设备, 2018, 38(9): 34-38. DOI: 10.16081/j.issn.1006-6047.2018.09.006.

ZHAO W Q, SHEN Z J, LI G. Anomaly detection for power consumption pattern based on deep learning[J]. Electric Power Automation Equipment, 2018, 38(9): 34-38. DOI: 10. 16081/j. issn. 1006-6047. 2018. 09. 006.

[9] BOLSTAD B M, IRIZARRY R A, ÁSTRAND M, et al. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias[J]. Bioinformatics, 2003, 19(2): 185-193. DOI: 10. 1093/bioinformatics/19. 2. 185.

[10] ZHANG S, ZHENG D Q, HU X C, et al. Bidirectional long short-term memory networks for relation classification[C]//29th Pacific Asia Conference on Language, Information and Computation. Shanghai, 2015: 73-78.

**作者简介:**  
谢禄江 男,(1974-),硕士,高级工程师. 研究方向为信息技术在电力行业的应用研究.  
段立 男,(1981-),硕士,高级工程师. 研究方向电力大数据分析应用.  
谭刚 男,(1969-),高级工程师. 研究方向为信息技术在电力行业的应用研究.  
蒋荣(通讯作者) 男,(1983-),硕士,高级工程师. 研究方向电力大数据及分析应用. E-mail: amy\_lj1220@163. com.  
钟淘淘 女,(1987-),工程师. 研究方向电力大数据及分析应用.  
刘美川 女,(1988-),经济师. 研究方向为大数据环境下的电力营销信息化建设等.  
雷洋 女,(1991-),助理工程师. 研究方向为电力营销.  
廖军 女,(1991-),硕士. 研究方向为大数据分析.