

Electricity Theft Detection Using Deep Bidirectional Recurrent Neural Network

Zhongtao CHEN*, De MENG*, Yufan ZHANG**, Tinglin XIN*, Ding XIAO**

*Information & Telecommunication Company, State Grid Jibei Electric Power Company Limited, Beijing, China

**School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China

18610816606@163.com, 308795810@qq.com, zhangyufan@bupt.edu.cn, 492510501@qq.com, dxiao@bupt.edu.cn

Abstract—Electricity theft causes significant harm to social and economic development. In the past few years, it has attracted much attention that electricity theft detection based on electricity consumption data can help to solve this problem. A major challenge is that there are no explicit features in electricity consumption records. However, the existing machine learning-based detection methods mainly suffer from the following two disadvantages. (1) Handcrafted features and shallow-architecture classifiers have poor detection accuracy. (2) Most methods consider electricity consumption as static and cannot capture both the internal time-series natures and external influence factors well. To overcome the above shortcomings, we propose a novel method called Electricity Theft Detection using Deep Bidirectional Recurrent Neural Network (ETD-DBRNN), which can capture the internal characteristics and the external correlation by learning the electricity consumption records and influence factors representation. Experiments on real-world datasets validate the effectiveness of our method.

Keywords—Electricity Theft Detection, AMI Data, Smart Grid, Deep Learning, Recurrent Neural Network

I. INTRODUCTION

The losses of power companies mainly come from two aspects [1], namely technical loss (TL) and non-technical loss (NTL). TL is naturally caused by internal electrical resistance or the affected components [2]; NTL is deliberately caused by the theft of electricity from malicious behavior. This misbehavior may bring about the revenue loss of power companies. For example, the loss from electricity theft cost the United States, United Kingdom, and Canada \$6 billion, \$173 million, and \$100 million annually. Therefore, *electricity theft detection* (ETD) is critical to prevent economic losses and security risks. However, conventional electricity-theft detection mainly relies on humanly checking or professional metering equipment, which is extremely time-consuming, expensive and inefficient.

In recent years, the advanced metering infrastructure (AMI) in smart grids has been widely deployed and it has greatly increased the sampling frequency of power consumption measurement data, thereby increasing the potential to detect power theft based on data-driven methods. However, traditional electricity theft detection methods suffer from the following two disadvantages.

First, existing efforts mainly considered electricity consumption as static and do not exploit the time-series nature and influence factors of the energy consumption data well. However, in the real world, electricity consumption behavior is often dynamic, which changes over time. The same electricity consumption appearing at different times may have a range of different impacts. For instance, low electricity consumption in summer or winter is more suspect than in spring because of higher electricity costs for refrigeration or heating equipment. In addition to the nature of time, electricity consumption is also closely related to the type of consumers. For example, the consumption of industrial customers often peaks during the weekdays, whereas residential customers consume most electricity on the weekends.

Second, the majority of existing works focus on the application of shallow machine learning techniques. The major challenge is that there are no explicit features in electricity consumption records. These shallow techniques including support vector machine (SVM) and other shallow neural networks based on handcrafted features (E.g. mean, standard deviation, maximum, minimum of consumption data), which have poor detection accuracy. On the other hand, deep learning-based neural networks are designed specifically to deal with huge amounts of data, and hence, capture the internal correlation and high-level features leading to improved detection performance.

As a deep learning architecture that learns sequence effectively, recurrent neural network (RNN) has a good performance when dealing with electricity consumption records with temporal properties. Bidirectional recurrent neural network (Bi-RNN) iterates in opposite directions respectively, which uses the forwards and backward information to enhance the overall learning effect of the sequence. Deep recurrent neural network (DRNN) is a neural network composed of multiple layers stacking on the basis of a single layer of RNN. Through increasing the number of RNN layers, the network can learn more abstract underlying features and strengthen the learning effect.

Therefore, we explore the use of recurrent neural networks for the detection task. We combine the advantages of DRNN and Bi-RNN, and propose a novel method called Electricity Theft Detection using Deep Bidirectional Recurrent Neural Network (ETD-DBRNN) for the electricity consumption theft

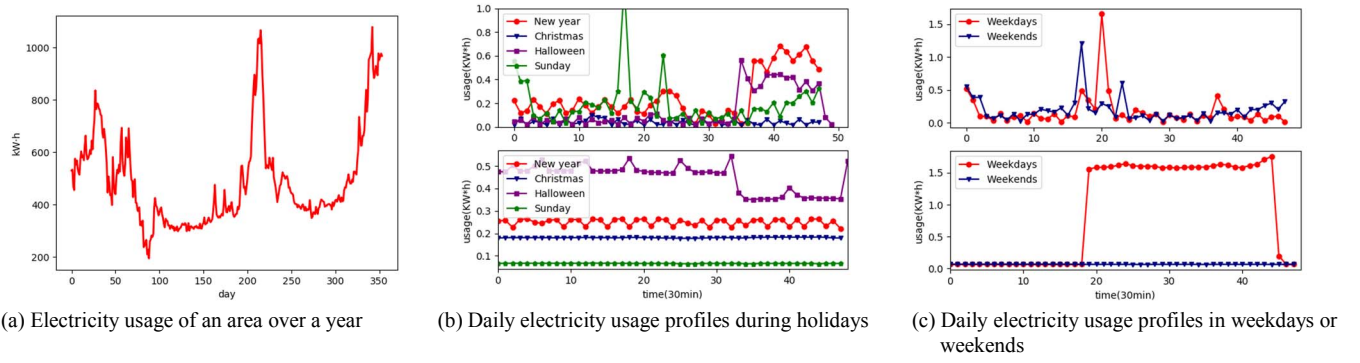


Figure 1. (a) shows the electricity consumption of an area over a year. (b) shows the daily electricity consumption during holidays, the top and bottom are residential and industrial consumers respectively. (c) shows the daily electricity consumption in weekdays or weekends, the top and bottom are residential and industrial consumers respectively.

detection task, which can automatically capture the internal characteristics and external correlation of the time-series data. In summary, our contributions come with as follows:

- We present a novel electricity theft detection method named ETD-DBRNN, which captures the deep information of the electricity consumption records by leaning the electricity consumption representation through a delicately designed recurrent neural network.
- We consider that the influence factors, including the type of consumers and dates, may impact on the detection and incorporate them into the records representation to improve the performance.
- We have conducted experiments on massive real-world electricity consumption dataset. Experimental results show that our ETD-DBRNN model outperforms other traditional approaches.

II. RELATED WORK

Electricity theft detection can be treated as a special case of anomaly detection or fraud detection. As a special time series, electricity consumption records have attracted much attention in the past few years. The daily electricity consumption patterns of users can be analyzed by using feature-based machine learning algorithms to establish classification models, such as linear regression (LR), artificial neural networks (ANN) [3], support vector machine (SVM) and so on. Many studies [4-7] classify the normal users and the electricity thieves based on SVM. In addition to SVM, a very recent study [8] utilized XGBoost, LightGBM, and CatBoost learning algorithms to detect NTL. However, these shallow machine learning methods cannot deal with high-dimensional data and extract deep features, and the low detection ratio is still not satisfactory enough for practice.

Recently, deep learning approaches have achieved huge success like image classification [9], speech recognition [10] and so on. Due to the ability to deal with complex information and automate feature extraction, which is work with smart meter data originating from the smart grids. A combination of grid features and external features based on neural network structures was proposed in [11], where the model was used for short-term load forecasting. A wide and deep CNN structure

was proposed in [12] to detect electricity theft in smart grid environments. Ref. [13] presents a deep architecture that exploits the temporal correlation based on RNN-based structure, which obtained a noticeable improvement performance than other shallow methods.

However, most existing related works in electricity consumption theft detection do not consider the internal and external features at the same time. Motivated by the ability to handle temporal and sequential information of RNN, we present a delicately designed RNN model called ETD-DBRNN combined with the advantages of deep layer and bidirectional layer, which can capture the electricity consumption records features in the high-level. Meanwhile, we utilize external information to strengthen the learned features and improve performance.

III. DATA ANALYSIS AND PRE-PROCESSING

In this part, we will analyze the electricity consumption data used in experiments, which provide the motivation of our model. And then, we explain the data pre-processing process according to the nature and characteristics of power data to make it suitable for the deep learning-based method.

A. Data Analysis

We conduct a preliminary analysis of real-world measurement data [14] released by Electric Ireland and Sustainable Energy Authority of Ireland (SEAI). This dataset contains half-hourly electricity consumption records (intervals of 30 minutes) of over 5000 Irish homes and businesses for a 533-days period from 2009 to 2010.

Figure 1(a) shows the electricity consumption of an area over a period of a year. We observe that the daily usage is changed over time, and it peaks in summer and winter because of the seasonality in time series. Therefore, the low electricity usage in winter or summer is more suspicious than the low electricity usage in spring or autumn.

Figure 1(b) and Figure 1(c) show the daily load profiles of two types of consumers including residential and industrial during holidays, weekdays or weekends. We observe that electricity consumption is periodic, and this periodicity is closely related to the type of consumers. In particular, for a

residential consumer, the daily load profiles are usually rich and varied due to the randomness of residential user behavior and power consumption is less affected by holidays. It peaks at almost the similar time in most days, which slightly changes during holidays. For an industrial consumer, the daily load profiles are usually single and stable because of the production activities, but the power consumption is greatly affected by holidays, that there is almost no consumption in holidays. Therefore, the low usage in holidays for residential or on weekdays for industrial is more suspicious.

In summary, the energy consumption records have internal time natures and correlate with external information such as the type of consumers and holidays. This motivates us to employ an RNN-based model for the proposed electricity theft detector since it can better deal with the time series nature of the energy data and capture the temporal correlation within the data. Meanwhile, we also observed that the daily consumption pattern is affected by calendar and type, which may strengthen the learning effect of our proposed model.

B. Data Pre-processing

The quality of data can be affected by various sensors, such as sensor power scale, noise level and so on. Therefore, it is necessary to apply appropriate pre-processing techniques to achieve a generalized performance.

Given a record data $\mathbf{x} = [x_1, x_2, \dots, x_T]$ with length T , x_t is defined as the energy consumption for an honest or malicious consumption record at time interval t . The basic steps of data pre-processing include missing value imputation, data smoothing, data normalization.

1) Missing Value Imputation

For a single missing data in a record, can be recovered by linear interpolation as follows:

$$f(x_t) = \frac{x_{t-1} + x_{t+1}}{2} \quad (1)$$

where x_t stands for the electricity usage of the time t .

2) Data Smoothing

In order to reduce the impact of noise points, a moving average is used to smooth the sequence. The K-term moving average of the sequence data at time t is:

$$f(x_t) = \frac{1}{K} \sum_{k=0}^K x_{t+k} \quad (2)$$

3) Data Normalization

Finally, data need to be normalized because the neural network is sensitive to the diverse data. One of the common methods for this scope is min-max normalization, which is computed as:

$$f(\mathbf{x}) = \frac{\mathbf{x} - x_{\min}}{x_{\max} - x_{\min}} \quad (3)$$

Where $x_{\min}$ and $x_{\max}$ represent the min and max values over a record $\mathbf{x}$, respectively.

IV. THE PROPOSED ETD-DBRNN

Suppose there are N_x historical electricity consumption records and each record has T intervals, we construct the electricity consumption records matrix $\mathbf{X} \in R^{N_x \times T}$ and the electricity theft matrix $\mathbf{Y} \in R^{N_x}$. Meanwhile, suppose there are N_o external conditions as influence factors $\mathbf{O} = [\mathbf{o}_1, \dots, \mathbf{o}_{N_o}]$. The problem of electricity theft detection is considered a discrete two-class classification task, which should assign each electricity consumption record $\mathbf{X}_i = [x_{i1}, x_{i2}, \dots, x_{iT}]$ into one of the predefined classes (abnormal or normal) as follows:

$$y_i = \begin{cases} 1, & \text{if record is abnormal;} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

In the following paper, we use electricity consumption records, external information sequence including days and types as time series $\mathbf{X}$, $\mathbf{o}^D$ and $\mathbf{o}^T$ respectively, and detect whether the electricity consumer records $\mathbf{X}$ have abnormal behaviors in the past time.

A. Framework Overview

The overall framework of ETD-DBRNN is illustrated in Figure 2, which is comprised of three key components: the input layer, representation layer, and output layer.

As we can see in Figure 2, to capture both internal and external information from the input layer, the electricity consumption series is fed into *Records Representation* using the recurrent neural network to learn the high-level representation. Meanwhile, external information including information of days and types is fed into *Factors Representation* using embedding and MLP to learn the representation of factors. Then we combine them by concatenate them to be the input of the final output layer. In the end, the prediction $\hat{y}_i$ for the given input is the output from upper-level MLP.

The overall formula in the model is as follows:

$$\hat{y}_i = f(\mathbf{X}_i, \mathbf{o}_j^D, \mathbf{o}_k^T | \mathbf{D}, \mathbf{T}, \mathbf{W}, \mathbf{U}, \mathbf{b}) \quad (5)$$

Where $\mathbf{X}_i$, $\mathbf{o}_j^D$, $\mathbf{o}_k^T$ denote the input vectors of the i -th record, j -th day, k -th type respectively and $\mathbf{D}$, $\mathbf{T}$, $\mathbf{W}$, $\mathbf{U}$, $\mathbf{b}$ denote the parameters of the model.

B. Input Layer

Consider a daily consumption record $\mathbf{X}_i = [x_{i1}, x_{i2}, \dots, x_{iT}]$ with length T as the left input of the model, where the element x_{it} is a record point, representing the sampling data of record i at time t . Then the input time series are reconstructed by a deep model based on a recurrent neural network.

Simultaneously, the right input layer accepts the j -th days and k -th types with one-hot encoding $\mathbf{o}_j^D \in R^{N_d}$ and $\mathbf{o}_k^T \in R^{N_t}$, where N_d and N_t are the number of days and types of users. It is reasonable because there will be a stronger similarity and temporal dependency among the same type of users or days in the latent space.

C. Representation Layer

1) Electricity Consumption Records Representation

We use GRU-based RNN to learn the sequential information of the electricity consumption records.

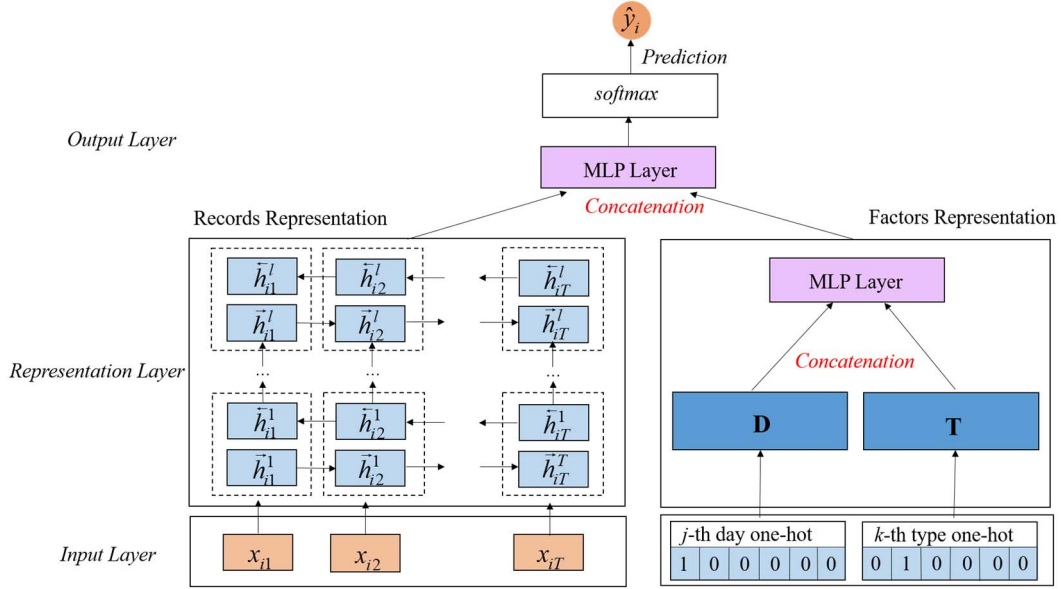


Figure 2. Electricity theft detection using deep bidirectional recurrent neural network framework

The gated recurrent unit (GRU) is a gating mechanism in a recurrent neural network to track the state of records without using separate memory cells. A GRU has two gates, reset gate r_t and update gate z_t , which together control how information is updated to the state. The GRU-based RNN function GRU is computed as follows:

At the time t , the GRU computes the new state as:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (6)$$

Where h_t is updated by linear interpolation from the previous state h_{t-1} and candidate state $\tilde{h}_t$. Formally, the update gate z_t is calculated as:

$$z_t = \sigma(\mathbf{W}_z x_t + \mathbf{U}_z h_{t-1} + \mathbf{b}_z) \quad (7)$$

Where x_t denotes the record vector at time t . The candidate state $\tilde{h}_t$ is updated according to the current input x_t and the previous hidden state h_{t-1} by a transition function:

$$\tilde{h}_t = \tanh(\mathbf{W}_h x_t + r_t \odot (\mathbf{U}_h h_{t-1}) + \mathbf{b}_h) \quad (8)$$

Where the reset gate r_t is calculated as:

$$r_t = \sigma(\mathbf{W}_r x_t + \mathbf{U}_r h_{t-1} + \mathbf{b}_r) \quad (9)$$

As shown in Figure 2, a deep bidirectional RNN (DBRNN) is constructed by stacking multiple layers. We stack multiple GRU hidden layers with forward and backward hidden layers, a deep bidirectional RNN based GRU is obtained.

At the first layer, the bidirectional GRU-based RNN contains the forward layer $\overrightarrow{GRU}$ which reads the record $\mathbf{X}_i$ from x_{i1} to x_{iT} and a backward layer $\overleftarrow{GRU}$ which reads the record $\mathbf{X}_i$ from x_{iT} to x_{i1} . The forward hidden state $\vec{h}_{it}$ and backward hidden state $\vec{h}_{it}$ are calculated as:

$$\vec{h}_{it}^1 = \overrightarrow{GRU}(x_{it}) \quad (10)$$

$$\vec{h}_{it}^1 = \overleftarrow{GRU}(x_{it}) \quad (11)$$

Then we get the candidate state $h_{it}^1 = [\vec{h}_{it}^1, \vec{h}_{it}^1]$ by concatenating them, which records the information of the whole electricity consumption sequence around x_{it} .

Finally, we get the output of the l -layer, the electricity consumption record embedding $\mathbf{E}_i^1 = [h_{i1}^l, \dots, h_{iT}^l]$ for $\mathbf{X}_i$, which will be fed into the output layer.

2) Influence Factors Representation

To enrich the representation of electricity consumption, *Factors Representation* receives the one-hot vector of days and users' type and output the embedding vectors.

As Figure 2 shows, the embedding matrix $\mathbf{D}_e$ and $\mathbf{T}_e$ map the sparse representations to dense vectors:

$$\mathbf{u} = \mathbf{D}_e \mathbf{o}_j^p \quad (12)$$

$$\mathbf{v} = \mathbf{T}_e \mathbf{o}_k^T \quad (13)$$

Where $\mathbf{u}$ and $\mathbf{v}$ are the embedding vectors of day j and type k . we concatenate them to be the input of a fully connected neural network, i.e. MLP (Multi-Layer Perceptron), the MLP with s layers as computed as follow:

$$\begin{aligned} \mathbf{a}_0 &= \begin{bmatrix} \mathbf{u} \\ \mathbf{v} \end{bmatrix} \\ \mathbf{a}_1 &= a(\mathbf{W}_1 \mathbf{a}_0 + \mathbf{b}_1) \\ &\vdots \\ \mathbf{a}_s &= a(\mathbf{W}_s \mathbf{a}_{s-1} + \mathbf{b}_s) \end{aligned} \quad (14)$$

Where a denotes the activation function for s layers, $\mathbf{a}_s$ is the out vector of factors representation layer for the record $\mathbf{X}_i$. Let $\mathbf{E}_i^2 = \mathbf{a}_s$, which will be fed into the output layer together with $\mathbf{E}_i^1$ from the records representation layer.

D. Output layer

In our proposed method, we design a deep bidirectional RNN to learn the electricity consumption records. As Fig.2 shows, the input of the deep neural network is electricity consumption sequence $\mathbf{X}_i$, which is encoded to an embedding $\mathbf{E}_i^1$ at layer l . The input of the factor embedding network is one-hot vectors of influence factors on $\mathbf{X}_i$, which is embedded to an embedding $\mathbf{E}_i^2$. After getting the embedding vectors $\mathbf{E}_i^1$ and $\mathbf{E}_i^2$, we concatenate them $\mathbf{E}_i = [\mathbf{E}_i^1, \mathbf{E}_i^2]$ to be a new input of the MLP unit with a non-linear activation layer, the anomalous probability $\hat{y}_i$ is computed as follows:

$$\hat{y}_i = \sigma(\mathbf{W}_o \cdot \mathbf{E}_i + \mathbf{b}_o) \quad (15)$$

Where $\mathbf{W}_o$ and $\mathbf{b}_o$ are the weight matrix of the output layer, σ is the softmax activation function.

After that, we use the negative log-likelihood of the correct labels as training loss:

$$L = - \sum_i [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)] \quad (16)$$

The model can be trained by minimizing the loss L between the predicted value $\hat{y}_i$ and the truth y_i .

V. EXPERIMENTS

A. Datasets

The real-world energy consumption data [14] from the Irish Smart Energy Trials are used to evaluate the performance of the proposed detector. This dataset contains three types of users, industrial, SME and other. Each type of user contains 535-day consumption records from July 2009 to December 2010. For a record in the datasets, which was sampled at 48 times a day. We random select records from three types of users for 365 days as our experimental datasets. The description of the datasets used is shown in Table 1.

Unfortunately, there are no labeled theft samples in the datasets. Studies on the data are based on the assumption that all users should belong to honest users because all users agreed to use their records for research purposes. In the real electricity consumption scene, the number of users stealing electricity is far less than normal users. Based on the fact, ref. [7] filtered most of the data as honest samples by the clustering algorithm, and generated six types of abnormal samples according to the daily consumption pattern.

In this paper, we follow this way to filter honest data and generate abnormal samples. After completing the necessary data preprocessing, 20% were randomly selected and modified to six types of abnormal samples.

TABLE 1. THE DATASETS

# of days N_d	365
# of types N_t	3
# of total records samples N_x	101200
# of daily sampling points T	48
# of residential samples	68355
# of SME samples	25410
# of other samples	7475

B. Baselines and Metrics

The key problem in ETD is how to learn electricity consumption records. In this paper, we use feature-based methods and deep learning-based methods as our baseline.

We use the following baselines:

- Machine learning-based methods: XGBoost, SVM.
- Deep learning-based methods: LSTM, GRU, Bi-LSTM, Bi-GRU. Where LSTM and GRU are two types of RNN cells.

Our ETD-DBRNN model with GRU-based cells noted as DBRNN-GRU. The DBRNN-GRU⁺ means it uses external

information including type and calendar to strengthen the effect of learning.

Given the prediction and its ground truth, a confusion matrix presented in Figure 3.

		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Figure 3. Confusion matrix

For binary classification, we compute the classic measure metrics of the model from the confusion matrix as follows:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (17)$$

$$P = \frac{TP}{TP + FP} \quad (18)$$

$$R = \frac{TP}{TP + FN} \quad (19)$$

$$F_1 = \frac{2PR}{P + R} \quad (20)$$

Where TP, TN, FP and FN are true positive, true negative, false positive and false negative in confusion matrix, respectively.

C. Experimental Setup

We conduct our experiments on a server with NVIDIA GeForce GTX 1080*2, and implement ETD-DBRNN based on Keras. We set the time step of each RNN layer to 48, the size of the hidden state of each RNN cell to 128 and the layers of deep RNN to 4. In the training, the batch size is 256, and the learning rate is 0.001. We use Adam to optimize our model.

D. Performance Comparison

The results of electricity theft detection are shown in Table 2. Note that the training data for all models use the raw records except the DBRNN-GRU⁺.

We can observe that in most cases, RNN-based methods achieve the best result. This indicates that the neural networks could capture the internal characteristics of anomalous samples and normal samples. Meanwhile, the sequential information of an electricity consumption pattern is critical to anomalous consumption detection. Compared to the neural network in one direction, the Bi-RNN based methods consider the positive and negative sequence effects in both directions and have a slight improvement.

The performance of the DBRNN and DBRNN⁺ are significantly better than other methods, which also shows that the deep structure can more effectively extract intent features. And we also can see that the external information is efficient to help to learn electricity consumption records embedding, which has greatly improved the recall rate index to 0.9974. This illustrates that the model is able to detect more electricity consumption theft records.

TABLE 2. PERFORMANCE COMPARISON WITH OTHER CONVENTIONAL SCHEMES

Methods	Acc	P	R	F ₁
SVM	0.8700	0.7459	0.3398	0.4669
XGBoost	0.9027	0.8697	0.4915	0.6281
RNN-LSTM	0.9379	0.8782	0.7296	0.7970
RNN-GRU	0.9401	0.9065	0.7158	0.7999
Bi-RNN-LSTM	0.9394	0.9161	0.7016	0.7946
Bi-RNN-GRU	0.9442	0.9150	0.7345	0.8149
DBRNN-GRU	0.9699	0.9586	0.8556	0.9042
DBRNN-GRU ⁺	0.9744	0.9570	0.9974	0.9209

VI. CONCLUSIONS

In this paper, we proposed Electricity Theft Detection using Deep Bidirectional Recurrent Neural Network (ETD-DBRNN). ETD-DBRNN learns the electricity consumption representation through a delicately designed recurrent neural network. Compared with traditional methods, ETD-DBRNN can capture the sequential information of the electricity consumption records and the internal characteristics between the normal and anomalous electricity consumption patterns. We also found that additional information can help improve the performance of detection.

The proposed model is one of the most typical examples which combines advanced deep learning techniques with traditional electric power. In the future, we will attempt to solve other electricity data mining tasks and make use of additional information to improve detection performance.

ACKNOWLEDGMENT

This paper is funded by the State Grid Jibei Power Co. Ltd. Science and Technology Project “Research on Data Fusion and Visualization Key Technologies Based on Full-service Unified Data Center” (52018E18006N).

REFERENCES

- [1] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni. Electricity theft: overview, issues, prevention and a smart meter based approach to control theft[J]. *Energy Policy*, 2011, 39(2): 1007-1015.
- [2] J. P. Navani, N. K. Sharma, and S. Sapra. Technical and nontechnical losses in power system and its economic consequence in Indian economy[J]. *International Journal of Electronics and Computer Science Engineering*, 2012, 1(2): 757-761.
- [3] DEPURU S S S R, WANG L, DEVABHAKTUNI V, et al. A hybrid neural network model and encoding technique for enhanced classification of energy consumption data[C]// 2011 IEEE Power and Energy Society General Meeting, July 24-29.
- [4] J. Naji, K. S. Yap, S. K. Tiong, et al. Nontechnical loss detection for metered customers in power utility using support vector machines[J]. *IEEE transactions on Power Delivery*, 2009, 25(2): 1162-1171.
- [5] S. S. S. R. Depuru, L. Wang, and V. Devabhaktuni. Support vector machine based data classification for detection of electricity theft[C]//2011 IEEE/PES Power Systems Conference and Exposition. IEEE, 2011: 1-8.
- [6] J. Naji, K. S. Yap, S. K. Tiong, et al. Improving SVM-based nontechnical loss detection in power utility using the fuzzy inference system[J]. *IEEE Transactions on power delivery*, 2011, 26(2): 1284-1285.
- [7] Jokar P., Arianpoo N., Leung, V.C. Electricity theft detection in AMI using customers' consumption patterns[J]. *IEEE Trans. Smart Grid*, 2015, 7(1): 216-226.

- [8] Coma-Puig, B., Carmona, J. Bridging the Gap between Energy Consumption and Distribution through Non-Technical Loss Detection. *Energies* 2019, 12, 1748.
- [9] Krizhevsky, A., Sutskever, I., Hinton, G.E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.
- [10] Hinton, G., Deng, L., Yu, D. et al. Deep neural networks for acoustic modeling in speech recognition[J]. *IEEE Signal. Process. Mag.* 2012, 29.
- [11] Chen P, Liu S, Shi C, et al. NeuCast: Seasonal Neural Forecast of Power Grid Time Series[C]//IJCAI. 2018: 3315-3321.
- [12] Zheng, Z., Yang, Y., Niu, X., et al. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. *IEEE Trans. Ind. Inform.* 2017, 14(4): 1606-1615.
- [13] Nabil M, Mahmoud M, Ismail M, et al. Deep Recurrent Electricity Theft Detection in AMI Networks with Evolutionary Hyper-Parameter Tuning[C]//2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). IEEE, 2019: 1002-1008.
- [14] Irish Social Science Data Archive. [Online]. Available: <http://www.ucd.ie/issda/data/commissionforenergyregulationcer/>



Zhongtao Chen (1986.1--). He is the supervisor and engineer at the Information and Telecommunication Company of the State Grid Jibei Electric Power Company Limited. He received his Ph.D. at the Chinese Academy of Sciences University. His research interests are communication networks analysis and data mining.



De Meng (1988.1--). He is an engineer at the Information and Telecommunication Company of the State Grid Jibei Electric Power Company Limited. His research interests are time series analysis and data mining.



Yufan Zhang (1996.10--). She is a Master Candidate at Beijing University of Posts and Telecommunications. Her research interests are deep learning and abnormal data detection.



Tinglin Xin (1989.8--). He is an engineer at the Information and Telecommunication Company of the State Grid Jibei Electric Power Company Limited. His research interests are deep learning and abnormal data detection.



Ding Xiao (1966.4--). He is a lecturer at Beijing University of Posts and Telecommunications. His research interests are software engineering and artificial intelligence.