International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Electricity Theft Detection in Power Grids with Deep Learning and Random Forests

Dr.U.Srinivas [1] and Mr.N.Satish Kumar[2]
[1]Assoc. Professor, Department of Information Technology, MREC (A), Hyderabad-500100
[2]Assistant Professor, Department of Information Technology, MREC (A), Hyderabad-500100

Abstract-As one of the major factors of the nontechnical losses (NTLs) in distribution networks, the electricity theft causes significant harm to power grids, which influences power supply quality and reduces operating profits. In order to help utility companies solve the problems of inefficient electricity inspection and irregular power consumption, a novel hybrid convolutional neural network-random forest (CNN-RF) model for automatic electricity theft detection is presented in this paper. In this model, a convolutional neural network (CNN) firstly is designed to learn the features between different hours of the day and different days from massive and varying smart meter data by the operations of convolution and down sampling. In addition, a dropout layer is added to retard the risk of over fitting, and the back propagation algorithm is applied to update network parameters in the training phase. And then, the random forest (RF) is trained based on the obtained features to detect whether the consumer steals electricity. To build the RF in the hybrid model, the grid search algorithm is adopted to determine optimal parameters. Finally, experiments are conducted based on real energy consumption data, and the results show that the proposed detection model outperforms other methods in terms of accuracy and efficiency.

Keywords: CNN, RF, NTL, hybrid model, electricity model.

1. INTRODUCTION

The loss of energy in electricity transmission and distribution is an important problem faced by power companies all over the world. The energy losses are usually classified into technical losses (TLs) and nontechnical losses (NTLs) [1]. The TL is inherent to the transportation of electricity, which is caused by internal actions in the power system components such as the transmission liner and transformers [2]; the NTL is defined as the difference between total losses and TLs, which is primarily caused by electricity theft. Actually, the electricity theft occurs mostly through physical attacks like line tapping, meter breaking, or meter reading tampering [3]. These electricity fraud behaviours may bring about the revenue loss of power companies. As an example, the losses caused by electricity theft are estimated as about $4.5 billion every year in the United States (US) [4]. And it is estimated that utility companies worldwide lose more than 20 billion every year in the form of electricity theft [5]. In addition, electricity theft behaviours can also affect the power system safety. For instance, the heavy load of electrical systems caused by electricity theft may lead to fires, which threaten the public safety. Therefore, accurate electricity theft detection is crucial for power grid safety and stableness.

With the implementation of the advanced metering infrastructure (AMI) in smart grids, power utilities obtained massive amounts of

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

electricity consumption data at a high frequency from smart meters, which is helpful for us to detect electricity theft [6, 7]. However, every coin has two sides; the AMI network opens the door for some new electricity theft attacks. These attacks in the AMI can be launched by various means such as digital tools and cyber attacks. The primary means of electricity theft detection include humanly examining unauthorized line diversions, comparing malicious meter records with the benign ones, and checking problematic equipment or hardware. However, these methods are extremely time-consuming and costly during full verification of all meters in a system. Besides, these manual approaches cannot avoid cyber attacks. In order to solve the problems mentioned above, many approaches have been put forward in the past years. These methods are mainly categorized into state-based, game-theory-based, and artificial-intelligence-based models [

The key idea of state-based detection [9–11] is based on special devices such as wireless sensors and distribution transformers [12]. These methods could detect electricity theft but rely on the real-time acquisition of system topology and additional physical measurements, which is sometimes unattainable. Game-based detection schemes formulate a game between electricity utility and theft, and then different distributions of normal and abnormal behaviours can be derived from the game equilibrium. As detailed in [13], they can achieve a low cost and reasonable result for reducing energy theft. Yet formulating the utility function of all players (e.g., distributors, regulators, and thieves) is still a challenge. Artificial-intelligence-based methods include machine learning and deep learning methods. Existing machine learning solutions can be further categorized into classification and clustering models, as is presented in [14–17]. Although aforementioned machine learning detection methods are innovative and remarkable, their performances are still not satisfactory enough for practice. For example, most of these approaches require manual feature extraction, which partly results from their limited ability to handle high-dimensional data. Indeed, traditional hand-designed features include the mean, standard deviation, maximum, and minimum of consumption data. The process of manual feature extraction is a tedious and time-consuming task and cannot capture the 2D features from smart meter data

Deep learning techniques for electricity theft detection are studied in [18], where the authors present a comparison between different deep learning architectures such as convolutional neural networks (CNNs), long-short-term memory (LSTM) recurrent neural networks (RNNs), and stacked autoencoders. However, the performance of the detectors is investigated using synthetic data, which does not allow a reliable assessment of the detector's performance compared with shallow architectures. Moreover, the authors in [19] proposed a deep neural network- (DNN-) based customer-specific detector that can efficiently thwart such cyber attacks. In recent years, the CNN has been applied to generate useful and discriminative features from raw data and has wide applications in different areas [20–22]. These applications motivate the CNN applied for feature extraction from high-resolution smart meter data in electricity theft detection. In [23], a wide and deep convolutional neural network (CNN) model was developed and applied to analyse the electricity theft in smart grids.

n a plain CNN, the softmax classifier layer is the same as a general single hidden layer feedforward

neural network (SLFN) and trained through the backpropagation algorithm [24]. On the one hand, the SLFN is likely to be overtrained leading to degradation of its generalization performance when it performs the backpropagation algorithm. On the other hand, the backpropagation algorithm is based on empirical risk minimization, which is sensitive to local minima of training errors. As mentioned above, because of the shortcoming of the softmax classifier, the CNN is not always optimal for classification, although it has shown great advantages in the feature extraction process. Therefore, it is urgent to find a better classifier which not only owns the similar ability as the softmax classifier but also can make full use of the obtained features. In most classifiers, the random forest (RF) classifier takes advantage of two powerful machine learning techniques including bagging and random feature selection which could overcome the limitation of the softmax classifier. Inspired by these particular works, a novel convolutional neural network-random forest (CNN-RF) model is adopted for electricity theft detection. The CNN is proposed to automatically capture various features of customers' consumption behaviours from smart meter data, which is one of the key factors in the success of the electricity theft detection model. To improve detection performance, the RF is used to replace the softmax classifier detecting the patterns of consumers based on extracted features. This model has been trained and tested with real data from all the customers of electricity utility in Ireland and London.

2. LITERATURE SURVEY

In this paper author is using combination of CNN (Convolution Neural Networks) and Random Forest to detect theft from electricity power grid as this theft will cause huge financial loss and disturbance in power supply. To efficiently detect theft from power grid author combining CNN and Random Forest Algorithms and after combining we are getting better prediction accuracy compare to normal algorithms. In power consumption if there is huge consumption in certain period then in dataset we will get value as 1 which indicates energy theft else we will have 0 as class label which means normal energy usage.

The main aim of the methodology described in this paper is to provide the utilities with a ranked list of their customers, according to their probability of having an anomaly in their electricity meter.

As shown in Figure 1, the electricity theft detection system is divided into three main stages as follows:(i)Data analysis and preprocess: to explain the reason of applying a CNN for feature extraction, we firstly analyse the factors that affect the behaviours of electricity consumers. For the data preprocess, we consider several tasks such as data cleaning (resolving outliers), missing value imputation, and data transformation (normalization).(ii)Generation of train and test datasets: to evaluate the performance of the methodology described in this paper, the preprocessed dataset is split into the train dataset and test dataset by the cross-validation algorithm. The train dataset is used to train the parameters of our model, whilst the test dataset is used to assess how well the model generalizes to new, unseen customer samples. Given that electricity theft consumers remarkably outnumber nonfraudulent ones, the imbalanced nature of the dataset can have a major negative impact on the performance of supervised machine learning methods. To reduce this bias, the synthetic minority oversampling technique (SMOT) algorithm is

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

used to make the number of electricity thefts and nonfraudulent consumers equal in the train dataset.(iii)Classification using the CNN-RF model: in the proposed CNN-RF model, the CNN firstly is designed to learn the features between different hours of the day and different days from massive and varying smart meter data by the operations of convolution and downsampling. And then, RF classification is trained based on the obtained features to detect whether the consumer steals electricity. Finally, the confusion matrix and receiver-operating characteristic (ROC) curves are used to evaluate the accuracy of the CNN-RF model on the test dataset.

3. METHODOLOGY

Agreeable means peoples who use words such as 'am, will have and this words also refers as ARTICLES or AUXILIARY VERBS' etc will come in this category. MRC dictionary contains all words of this categories and by applying this dictionary on user's tweets we can predict person category as agreeable.

Neuroticism means peoples in this category is consider as sentiment or emotion, peoples who use words such as 'ugly, nasty, sad' etc will come under this category. By looking for such words in tweets we can predict score of this category.

Extroversion means peoples of this category are friendly and person who has many number of friends or followers or following in twitter profile will comes under this category.

Conscientious means peoples who express hard working ideas in their post will come under this category.

So by analysing above 5 features O (openness), C (Conscientious), E (Extroversion), A (Agreeable), N (Neuroticism) from twitter profile and post we can predict personality of a person.

We will find average of each feature from tweets and then apply Pearson Correlation formula to get score for all five features. If score > 0.1 for any feature then person belongs to that category. If person has 0.1 value for more than 1 features then that person personality belongs to that many categories. For example same person can be predicted as openness and conscientious etc.

3. RESULTS

All features values we will apply using SVM, Random Forest, Naïve Bayes & Logistic Regression algorithms to calculate accuracy of dataset and algorithms.

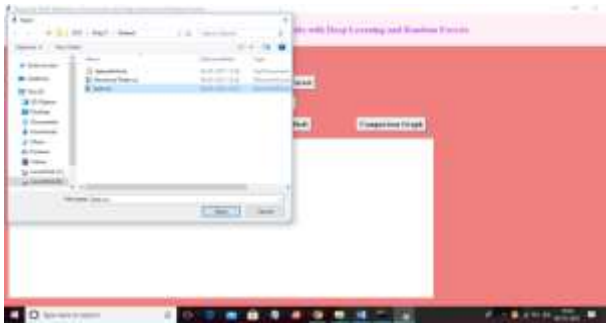
In above screen with CNN-RF we got 100% accuracy and now click on 'CNN with SVM' button to train dataset with CNN and SVM

In above screen with CNN-SVM we got 99% accuracy and now click on 'Run Random Forest' button to train alone RF on dataset

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

In above screen with alone Random Forest we got 94% accuracy and now click on 'Run SVM Algorithm' button to train alone SVM with above dataset

In above screen with alone SVM we got 96% accuracy and now click on 'Predict Electricity Theft' button to upload test data



In above screen selecting and uploading 'test.csv' file and then click on 'Open' button to load test data and to get below prediction result

In above screen in square brackets we can see test data and after square bracket we can see prediction result as 'record detected as ENERGY THEFT' or 'record NOT detected as ENERGY THEFT'. Now click on 'Comparison Graph' button to get below graph

In above graph x-axis represents algorithm names and y-axis represents precision, recall, FSCORE and Accuracy for each algorithm and in all algorithms CNN-RF is giving 100% accuracy.

5. CONCLUSION

In this paper, a novel CNN-RF model is presented to detect electricity theft. In this model, the CNN is similar to an automatic feature extractor in investigating smart meter data and the RF is the output classifier. Because a large number of parameters must be optimized that increase the risk of overfitting, a fully connected layer with a dropout rate of 0.4 is designed during the training phase. In addition, the SMOT algorithm is adopted to overcome the problem of

data imbalance. Some machine learning and deep learning methods such as SVM, RF, GBDT, and LR are applied to the same problem as a benchmark, and all those methods have been conducted on SEAI and LCL datasets. The results indicate that the proposed CNN-RF model is quite a promising classification method in the electricity theft detection field because of two properties: The first is that features can be automatically extracted by the hybrid model, while the success of most other traditional classifiers relies largely on the retrieval of good hand-designed features which is a laborious and time-consuming task. The second lies in that the hybrid model combines the advantages of the RF and CNN, as both are the most popular and successful classifiers in the electricity theft detection field. Since the detection of electricity theft affects the privacy of consumers, the future work will focus on investigating how the granularity and duration of smart meter data might affect this privacy. Extending the proposed hybrid CNN-RF model to other applications (e.g., load forecasting) is a task worth investigating.

REFERANCES

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," *Energy Policy*, vol. 39, no. 2, pp. 1007–1015, 2011.View at: Publisher Site | Google Scholar
2. J. P. Navani, N. K. Sharma, and S. Sapra, "Technical and non-technical losses in power system and its economic consequence in Indian economy," *International Journal of Electronics and Computer Science Engineering*, vol. 1, no. 2, pp. 757–761, 2012.View at: Google Scholar
3. S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, "A multi-sensor energy theft detection framework for advanced metering infrastructures," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1319–1330, 2013.View at: Publisher Site | Google Scholar
4. P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Security & Privacy Magazine*, vol. 7, no. 3, pp. 75–77, 2009.View at: Publisher Site | Google Scholar
5. T. B. Smith, "Electricity theft: a comparative analysis," *Energy Policy*, vol. 32, no. 1, pp. 2067–2076, 2004.View at: Publisher Site | Google Scholar

6. J. I. Guerrero, C. León, I. Monedero, F. Biscarri, and J. Biscarri, "Improving knowledge-based systems with statistical techniques, text mining, and neural networks for non-technical loss detection," *Knowledge-Based Systems*, vol. 71, no. 4, pp. 376–388, 2014.View at: Publisher Site | Google Scholar
7. C. C. O. Ramos, A. N. Souza, G. Chiachia, A. X. Falcão, and J. P. Papa, "A novel algorithm for feature selection using harmony search and its application for non-technical losses detection," *Computers & Electrical Engineering*, vol. 37, no. 6, pp. 886–894, 2011.View at: Publisher Site | Google Scholar
8. P. Glauner, J. A. Meira, P. Valtchev, R. State, and F. Bettinger, "The challenge of non-technical loss detection using artificial intelligence: a surveyficial intelligence: a survey," *International Journal of Computational Intelligence Systems*, vol. 10, no. 1, pp. 760–775, 2017.View at: Publisher Site | Google Scholar
9. S.-C. Huang, Y.-L. Lo, and C.-N. Lu, "Non-technical loss detection using state estimation and analysis of variance," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2959–2966, 2013.View at: Publisher Site | Google Scholar
10. O. Rahmati, H. R. Pourghasemi, and A. M. Melesse, "Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: a case study at Mehran region, Iran," *CATENA*, vol. 137, pp. 360–372, 2016.View at: Publisher Site | Google Scholar