Journal of Soft Computing Paradigm (ISSN: 2582-2640)
www.irojournals.com/jscp/

Identification of Electricity Threat and Performance Analysis using LSTM and RUSBoost Methodology

Joy Iong-Zong Chen[1], Lu-Tsou Yeh[2]

[1]Professor, Department of Communication Engineering, Da-Yeh University, Chang-Hua, Taiwan
[2]Professor, Department of Electrical Engineering, Da-Yeh University, Chang-Hua, Taiwan

E-mail: [1]jchen@mail.dyu.edu.tw

Abstract

In power systems, electrical losses can be categorized into two types, namely, Technical Losses (TLs) and Non-Technical Losses (NTLs). It has been identified that NTL is more hazardous when compared to TL, primarily due to the factors such as billing errors, faulty meters, electricity theft etc. This proves to be crucial in the power system and will result in heavy financial loss for the utility companies. To identify theft, both academia and industry, use a mechanism known as Electricity Theft Detection (ETD). However, ETD is not used efficiently because of handling high-dimensional data, overfitting issues and imbalanced data. Hence, in this paper, a means of addressing this issue using Random Under-Sampling Boosting (RUSBoost) technique and Long Short-Term Memory (LSTM) technique is proposed. Here, parameter optimization is performed using RUSBoost and abnormal electricity patterns are detected by LSTM technique. Electricity data are pre-processed in the proposed methodology, using interpolation and normalization methods. The data thus obtained are then sent to the LSTM module where feature extraction takes place. These features are then classified using RUSBoost algorithm. Based on the output simulated, it is identified that this methodology addresses several issues such as handling and overfitting of massive time series data and data imbalancing. Moreover, this technique also proves to be more efficient than several other methodologies such as Logistic Regression (LR), Convolutional Neural Network (CNN) and Support Vector Machine (SVM). A comparison is also drawn, taking into consideration the parameters such as Receiver operating characteristics, recall, precision and F1-score.

Keywords: Parameter Tuning, Imbalanced data, Random under sampling, Smart meter, Electricity theft, non-technical losses, Long Short-Term Memory, RUSBoost

DOI: https://doi.org/10.36548/ jscp.2021.4.001
Received: 16.10.2021, received in revised form: 24.10.2021, accepted: 10.11.2021, published: 22.11.2021

1. Introduction

The amount of energy spent which is not billed by the consumer is known as Electricity theft. This results in heavy revenue losses that will affect electric utility companies. It has been observed that electricity theft accounts for several crores of revenue loss for electric utility companies across the globe. This in turn will affect the development of the country [1]. In India, over 5 billion rupees is lost annually due to electricity theft while Pakistan suffers a loss of 0.9 billion rupees a year. Companies with strong economies such as the US and UK are also subjected to electricity theft building up to $7 billion in the U.S. and £175 million per annum in the UK. Moreover, electricity theft also results in voltage imbalance which will affect the operating transformers by overloading them. Similarly, electricity theft will also cause the increase in electricity price, decrease in energy efficiency, decrease in revenue, increase in inflation rate and increase in unemployment, thereby affecting the economic growth of the country [2]. NTL occurs due to unregistered connections [3], direct hooking [4], meter tampering [5] and meter modifications [6]. It is possible to identify electricity theft manually with the help of on-field inspections.

Recovery of NTL is performed by the inspection team with the help of the readings observed from the meters, to identify the faulty meters. However, this process of inspection will consume a lot of time and will also require the use of a professional team. Moreover, installation of hardware to identify electricity thefts is also essential when using this methodology. In recent years, traditional grids have been replaced and evolved with the introduction of smart meters that are data-driven and prove to be quite efficient in energy management [7]. Deep learning algorithms and data driven techniques have been introduced to replace the traditional methodologies. The advantages of these algorithms are their adaptability and efficiency in number applications. They study the abnormal electricity consumption and use this data to identify the amount of energy stolen and ways of predicting electricity theft. However, the algorithms developed so far have several shortcomings which are experimented and researched by industrialists and researchers [8].

In this proposed work, a bat-based random under sampling boosting and long short term memory are incorporated. Here, the classifier biasing nature is avoided and imbalance problems are addressed with the help of RUSBoost [9] and the sequential model with long short term memory is used with the LSTM model. To validate the efficiency of the proposed work, it is compared with several other algorithms [10]. The following are the contributions of this paper:

- To memorize and extract features in a better manner, LSTM block [11] is used. This block efficiently identifies and extracts information regarding the theft of electricity.

- Normalization and interpolation methodologies are used for data pre-processing. These methods enable collection of data and determination of missing data [12].

- The imbalance problems [13]-[16] are handled with the help of RUSBoost algorithm and it further executes in a more efficient manner when compared to previously existing algorithms. This algorithm executes in two steps. The first step involves data under-sampling and the next is the final classification prediction with Adaboost. This algorithm has the advantage of learning from previous mistakes which in turn improves the efficiency of the output.

- A comparison analysis is drawn between receiver operating characteristics, recall, precision and F1 score to calculate the model's accuracy [17].

2. Related Works

The proposed work can be classified into three categories namely Machine Learning (ML) [18], game theory [19] and state-based solutions [20]. Here, the machine learning based solutions make use of the data of energy consumed to determine the actual usage of the customer, resulting in identification of the thief. On the other hand the game theory solutions, consider the happenings between electric utilities and the energy thief, to be a game [21]. Based on the energy difference between the energy consumption distribution and the outcome of the game, the thief is identified. This is a low cost solution that proves to be effective. However, the drawback is that it requires more time to analyze [22] and identify the theft. On the other hand, the state-based solution suggests that the distribution transformers and design of specific metering devices [23] should be focused to identify electricity theft. Though this methodology proves to be effective in identifying electricity theft, they still require additional hardware tools like distribution transformers and meter sensors that are economically pricey. These methodologies can also be categorized into supervised, semi-supervised and clustering techniques. Table 1 gives an overview of the various machine learning techniques that exist, their limitations and the means of validation.

Table 1. Existing Machine Learning Techniques and their Limitations

Methodology	Dataset Used	Limitations	Validation Parameters	Purpose
LSTM, GMM [24]	Numenta Anomaly Benchmark	Not Robust	F1 score	LSTM internal architecture is advanced that positively influence the performance
Auto-encoder [25]	Hong Kong data	Overfitting	Accuracy	When used in commercial buildings, anomaly detection is improved with Auto-encoder
SMOTE, CNN [26] and RF	EISA	Increased execution duration	F1 score	Using RF in the last layer, it is possible to remove local optima
LSTN, SMOTE [27] and CNN	SGCC	Over fitting	F1 score	Fraudulent customer detection improves
Wide and deep CNN [28]	SGCC	Imbalance of Data	MAP, AUC	Using extracted global and local features, captures electricity theft
MLP, LSTM [29]	Endesa	Imbalance of data	PR AUC, ROC	Combination of sequential data and auxiliary information to identify electricity theft
RUSBoost, MODWPT [30]	Honduras	No parameter tuning	F1 score, MCC	More efficient in identifying NTL

3. Proposed Work

The proposed system model for ETD comprises of three stages of operation namely: data preprocessing stage, feature extraction stage and feature refinement stage. A block diagram of the stages and the operation performed by the proposed work is given in Fig.1.

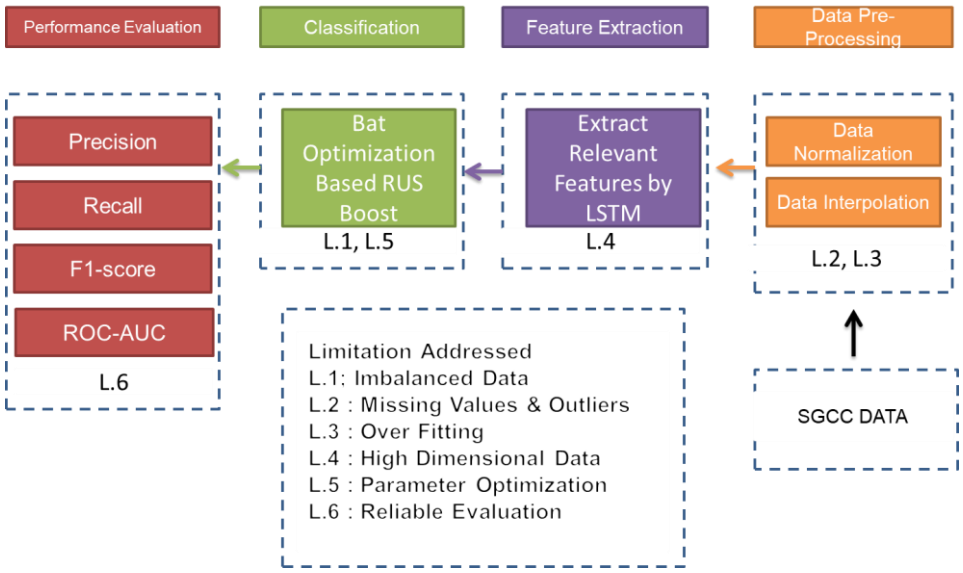

Figure 1. Block Diagram of the Proposed Model

3.1 Data Preprocessing

In general, the actual electricity consumption data are not perfect and have many missing values which can be rectified with the help of data preprocessing. This step is crucial in improving the performance of the classifier primarily because the quality of input data will have a major impact on the ML algorithm. Data gathered from the SGCC will hold many erroneous and missing data that is primarily caused by unreliable transmission and faulty measuring instruments. When datasets go amiss, it becomes difficult for the classifier to find the fraudulent consumers. Similarly, scattered data will also lead to difficulty in theft identification and increases the time taken for execution. Hence two stages are executed under data preprocessing namely data interpolation and data normalization. For data normalization, the inputs are used with a common scale in order to maintain a common range. Normalization takes place with the help of Equation (1)

$$B' = \frac{B - Min\ (B)}{Max\ (B) - Min\ (A)} \times ((C - A) + A) \tag{1}$$

where B' is the normalized value and C and A are the minimum and maximum values, respectively. This ensures data analysis and decreases the execution time. When B is maximum, B'=1. It indicates that 0 is mapped to minimum value while 1 is mapped to maximum value.

Similarly data interpolation is carried out using the following equation (2)

$$f(y_i) = \begin{cases} \frac{y_{i+1}+y_{i-1}}{2} & if\ y_i \in NaN, y_{i+1}\ and\ y_{i-1} \notin NaN \\ 0, & if\ y_i \in NaN, y_{i+1}\ and\ y_{i-1} \notin NaN \\ y_i & if\ y_i \notin NaN \end{cases} \quad (2)$$

where NaN denotes non-numeric value, y_i is the attribute of electricity consumption data and y_{i+1} and y_{i-1} are non-numeric values in dataset.

3.2 Feature Extraction

The next stage in this process is feature extraction wherein an LSTM module is used to extract the essential information and refine them. Due to the enormous data collection, it is not possible to use recurrent neural network. However a variant of RNN, LSTM, which is capable of handling gradient exploding and vanishing is used for this purpose. LSTN also shows positive aspect in the classification of large time series data and capturing of temporal correlations.

3.3 Bat Algorithm

Parameter tuning is the primary concern as far as accuracy of classification is considered. In this methodology, a bat algorithm is used to pick the best parameter value for RUSBoost. This methodology is based on the echolocation characteristics of the bats.

4. Results and Discussion

Support Vector Machine (SV) and Synthetic Minority Over-sampling Technique (SMOTE) are existing methodologies of machine learning. In this section, a comparison is drawn between these existing methodologies and the proposed work. In SVM, the complexity of training is influenced by the input data and Fig.2 and Fig.3 show the failure of SVM due to its unsuccessful attempt at identifying the fraud users. On the other hand, the proposed work shows significant accuracy in identifying the theft and the fraud users in all respects of the performance metrics.

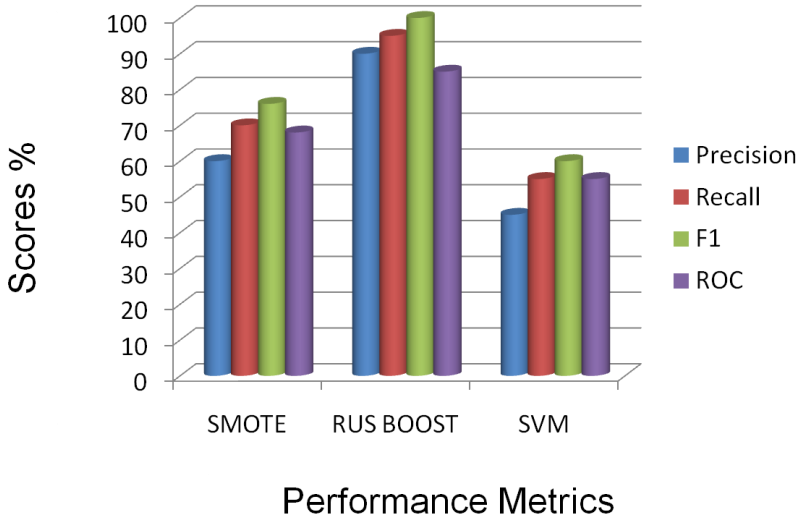

Figure 2. Comparison of Unbalanced and Balanced Data with the Proposed Methodology

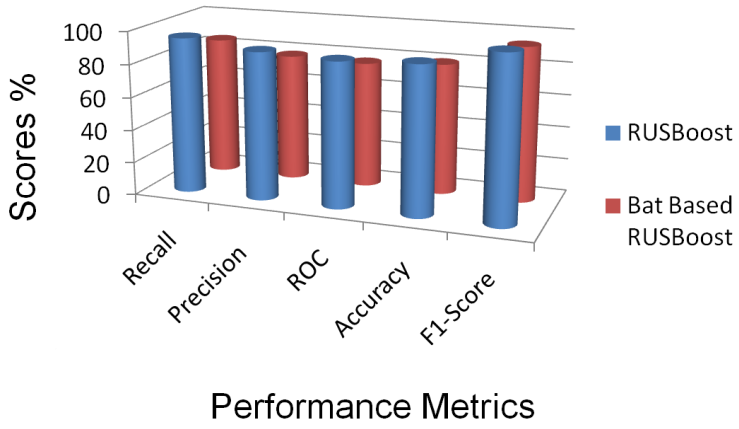

Figure 3. Performance comparison of parameter tuning

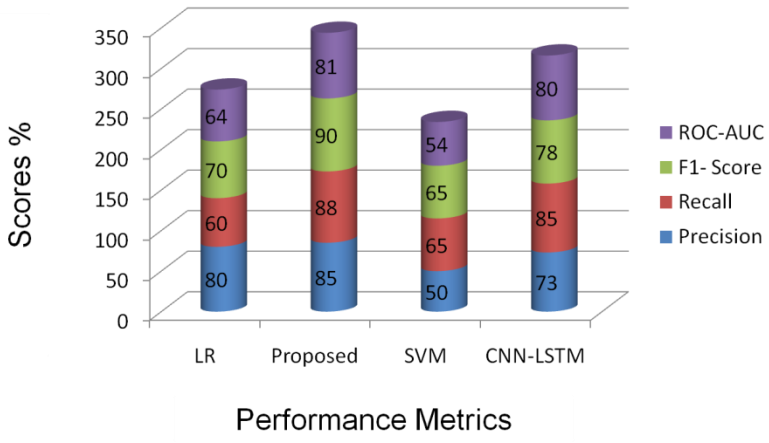

Figure 4. Comparison of Performance

To validate the proposed model and its performance, the output is compared with other schemes that are used in many applications. The performance metrics like ROC curve, F1-score, recall and precision are taken into consideration in Fig.4. Benchmark models like CNN-LSTM, SVM and Logistic Regression (LR) are used for the purpose of comparison.

5. Conclusion

In this paper, a novel electricity theft detection is introduced and evaluated with the help of real-time information. Normalization and Interpolation methods are used on the electricity data to pre-process and get rid of undefined and null values. This is followed by refinement of features using LSTM methodology that can extract the required information from the pre-processing information. The last step in this process is the use of the RUSBoost method which is used to categorize the data based on the type of customers as honest and dishonest, thereby providing a good balance. The parameters considered are optimized with the help of bat algorithm in RUSBoost method. The simulated output is then compared with other methodologies such as CNN-LSTM, LR and SVM techniques. Based on the evaluation results, it is found that the simulation obtained from the proposed model is more efficient in terms of overfitting, parameter optimization and handling of imbalance data. Based on the performance metrics analyzed it is found that the proposed work attains high precision, recall, F1-score and ROC-AUC. However, it is also observed that the output is highly sensitive to any deviation in input data and will require attention as the part of future work.

References

[1] Vijayakumar, T. "Comparative study of capsule neural network in various applications." Journal of Artificial Intelligence 1, no. 01 (2019): 19-27.

[2] Huckle, S., Bhattacharya, R., White, M., &Beloff, N. (2016). Internet of things, blockchain and shared economy applications. Procedia computer science, 98, 461-466.

[3] Patil, Prachu J., Ritika V. Zalke, Kalyani R. Tumasare, Bhavana A. Shiwankar, Shivani R. Singh, and Shailesh Sakhare. "IoT Protocol for Accident Spotting with Medical Facility." Journal of Artificial Intelligence 3, no. 02 (2021): 140-150.

[4] Casado-Vara, R., Prieto, J., De la Prieta, F., &Corchado, J. M. (2018). How blockchain improves the supply chain: Case study alimentary supply chain. Procedia computer science, 134, 393-398.

[5] Manoharan, J. Samuel. "A Novel User Layer Cloud Security Model based on Chaotic Arnold Transformation using Fingerprint Biometric Traits." Journal of Innovative Image Processing (JIIP) 3, no. 01 (2021): 36-51.

[6] Kumar, N. M., &Mallick, P. K. (2018). Blockchain technology for security issues and challenges in IoT. Procedia Computer Science, 132, 1815-1823.

[7] Jacob, I. Jeena, and P. Ebby Darney. "Artificial Bee Colony Optimization Algorithm for Enhancing Routing in Wireless Networks." Journal of Artificial Intelligence 3, no. 01 (2021): 62-71.

[8] Sikorski, J. J., Haughton, J., & Kraft, M. (2017). Blockchain technology in the chemical industry: Machine-to-machine electricity market. Applied energy, 195, 234-246.

[9] Mugunthan, S., and T. Vijayakumar. "Review on IoT based smart grid architecture implementations." j Electric Eng Autom 1, no. 1 (2019): 12-20.

[10] Zhang, S., & Lee, J. H. (2020). Analysis of the main consensus protocols of blockchain. ICT express, 6(2), 93-97.

[11] Smys, S., Haoxiang Wang, and Abul Basar. "5G Network Simulation in Smart Cities using Neural Network Algorithm." Journal of Artificial Intelligence 3, no. 01 (2021): 43-52.

[12] Gordon, W. J., & Catalini, C. (2018). Blockchain technology for healthcare: facilitating the transition to patient-driven interoperability. Computational and structural biotechnology journal, 16, 224-230.

[13] Shakya, Subarana. "An efficient security framework for data migration in a cloud computing environment." Journal of Artificial Intelligence 1, no. 01 (2019): 45-53.

[14] Dasgupta, D., Shrein, J. M., & Gupta, K. D. (2019). A survey of blockchain from security perspective. Journal of Banking and Financial Technology, 3(1), 1-17.

[15] Raj, Jennifer S. "Optimized Mobile Edge Computing Framework for IoT based Medical Sensor Network Nodes." Journal of Ubiquitous Computing and Communication Technologies (UCCT) 3, no. 01 (2021): 33-42.

[16] Beck, R., Avital, M., Rossi, M., & Thatcher, J. B. (2017). Blockchain technology in business and information systems research.

[17] Bashar, Abul. "Survey on evolving deep learning neural network architectures." Journal of Artificial Intelligence 1, no. 02 (2019): 73-82.

[18] Banotra, A., Sharma, J. S., Gupta, S., Gupta, S. K., & Rashid, M. (2021). Use of blockchain and internet of things for securing data in healthcare systems. In Multimedia Security (pp. 255-267). Springer, Singapore.

[19] Wang, Haoxiang. "IoT based Clinical Sensor Data Management and Transfer using Blockchain Technology." Journal of ISMAC 2, no. 03 (2020): 154-159.

[20] Namasudra, S., Deka, G. C., Johri, P., Hosseinpour, M., &Gandomi, A. H. (2021). The revolution of blockchain: State-of-the-art and research challenges. Archives of Computational Methods in Engineering, 28(3), 1497-1515.

[21] Upadhyay, Hemant, YogeshKamat, ShubhamPhansekar, and Varsha Hole. "User Engagement Recognition Using Transfer Learning and Multi-task Classification." In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 411-420. Springer Singapore, 2021.

[22] Shirley, D. R. A. (2014, July). Systematic diagnosis of power switches. In 2014 International Conference on Embedded Systems (ICES) (pp. 32-34). IEEE.

[23] Varsha, Viswanathan, and C. N. Sminesh. "QoS Aware Multi Mapping Technology in SD-WAN." In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 421-433. Springer Singapore, 2021.

[24] Shirley, D. R. A., Amruthavarshni, R. B., Durainathan, A., &Karthika, M. P. (2021, May). QR-Based inventory management system (QR-IMS) of passenger luggage using website. In 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1180-1185). IEEE.

[25] Dandagi, Vidya S., and Nandini Sidnal. "Auto-Completion of Queries." In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 435-446. Springer Singapore, 2021.

[26] Sharples, M., & Domingue, J. (2016, September). The blockchain and kudos: A distributed system for educational record, reputation and reward. In European conference on technology enhanced learning (pp. 490-496). Springer, Cham.

[27] Mishra, Zishani, T. Prashanth, N. Sanjay, Jagrati Gupta, and Amit Jain. "Design of CMOS Active Inductors for RFIC Applications: A Review." In Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 447-456. Springer Singapore, 2021.

[28] Seebacher, S., &Schüritz, R. (2017, May). Blockchain technology as an enabler of service systems: A structured literature review. In International Conference on Exploring Services Science (pp. 12-23). Springer, Cham.

[29] Subbulakshmi, S., K. Ramar, Anvy Elsa Saji, and GeethuChandran. "Optimized Web Service Composition Using Evolutionary Computation Techniques." In Intelligent Data

Communication Technologies and Internet of Things: Proceedings of ICICI 2020, pp. 457-470. Springer Singapore, 2021.

[30] Qiu, H., Qiu, M., Memmi, G., Ming, Z., & Liu, M. (2018, December). A dynamic scalable blockchain based communication architecture for iot. In International Conference on Smart Blockchain (pp. 159-166). Springer, Cham. Author's biography

Author's biography

Joy Iong-Zong Chen is currently a full professor in the Department of Communication Engineering Dayeh University at Changhua Taiwan. Prior to joining Dayeh University, he worked at the Control Data Company (Taiwan) as a technical manager from Sep. 1985 toSep. 1996. His research interests include wireless communications, spread spectrum technical, OFDM systems, and wireless sensor networks. He has published a large number of SCI Journal papers on the issues addressed by the physical layer for wireless communication systems. Moreover, he also majors in developing some applications of the IOT (Internet of Thing) techniques and Dr. Joy I.-Z. Chen owned some patents authorized by the Taiwan Intellectual Property Office (TIPO).

Lu-Tsou Yeh works as a Professor in the Department of Electrical Engineering, Da-Yeh University, Chang-Hua, Taiwan. His major areas of research are semiconductor materials, computer science, nano electronics, object/web technologies, microelectronics, quantum electronics, VLSI, electronic system design, IT integrated manufacturing, fabrication, and analysis which remains as the backbone for developing next generation electronic devices and information technology applications.