Deep learning for combating energy theft: Modeling and analysis of Hybrid LSTM

Hongxin Gao

Technical Report

RHUL–ISG–2022–6

11 April 2022

Information Security Group
Royal Holloway University of London
Egham, Surrey, TW20 0EX
United Kingdom

Executive Summary

In the energy industry, power theft has been a constant concern for many countries. With the development of technology, the means of electricity theft also reflects the characteristics of advanced and concealed. However, the emergence of artificial intelligence technology has also elevated the combat against electricity theft to a new level. The classical machine learning has been proven to be effective in electricity theft detection (ETD). Instead of further corroborating the feasibility of them, this project further explores the usability of advanced models in the emerging field of deep learning. Recent academic journals and research results show that advanced models and algorithms based on convolutional neural network (CNN), long short-term memory (LSTM), and generative adversarial network (GAN), among others, are also effective and even more flexible and efficient in ETD.

This project uses a data set of electricity consumption of real users, which contains data and labels of normal users as well as electricity theft users, i.e., a supervised learning approach is used to deal with the binary classification issue. The performance of the target model on the unseen data set is evaluated by training the model and algorithm, i.e., the performance of the model in identifying normal electricity users and electricity theft users. During the project, some advanced and effective deep learning techniques are also well applied and show good results. In terms of data pre-processing, techniques such as k-nearest neighbors (KNN) imputation processing for missing values, interquartile range (IQR) processing for outliers and borderline-synthetic minority oversampling technique (Borderline-SMOTE) processing for imbalance classification were successfully applied to the project. Also, the clear visualization analysis provides a good basis for modeling. Three models were successfully built, with multilayer perceptron (MLP) as the baseline model, CNN & LSTM (CNN-LSTM) as a comparable model and convolutional LSTM (ConvLSTM) as a novel model. The model convergence is accelerated by optimizing hyperparameters such as dropout and learning rate. And more comprehensive metrics are used on the unseen data to evaluate the feasibility, accuracy, and robustness of the model in identifying electricity theft users.

In this project, ConvLSTM outperformed the other models with accuracy, loss, precision, recall, F1-score, Cohen's kappa, receiver operating characteristic-area under the curve (ROC-AUC) and precision-recall-AUC (PR-AUC) of 0.984, 0.089, 0.984, 0.985, 0.984, 0.969, 0.993, 0.991 respectively. In addition, ConvLSTM supports multi-dimensional electricity data input for better extraction of time series features, and batch normalization technology supports direct transformation of raw electricity data in model compilation without tedious and time-consuming data pre-transformation. This also demonstrates that ConvLSTM also shows a lot of room for improvement in terms of flexibility of model architecture adjustment and efficiency of data processing. This can also better improve the timeliness of power companies in combating electricity theft, and enable them to adjust their detection deployment in a timely manner in an environment where theft methods are constantly changing.

A journal version of this project report is pending submission to International Transactions on Electrical Energy Systems for review.