Detection of Non-Technical Losses using Advanced Metering Infrastructure and Deep Recurrent Neural Networks

Soham Chatterjee*, Vaidheeswaran Archana*, Karthik Suresh*, Rohit Saha†, Raghav Gupta† and Fenil Doshi†
*Department of Electrical and Electronics Engineering
SRM University, Chennai, India
Email: 96soham96@gmail.com
†Department of Computer Science and Engineering
SRM University, Chennai, India
Email: r.saha1997@gmail.com

***Abstract*—Non-Technical Losses including power theft is a major problem faced by power utilities today. In India, this accounts for roughly 50% of the total loss every year. Non-Technical losses can be infidelity at the consumer end, unethical tapping of transmission lines and hacking or tampering of energy meters. In this paper, a solution to power theft using advanced metering infrastructure and intelligent algorithms has been proposed. The proposed solution profiles users based on their half hour wise power consumption data and locality. Irregularities in power usages are detected using a state-of-the-art artificially intelligent algorithm: Recurrent Neural Networks. The algorithm uses Long Short Term Memory (LSTM) units to process sequential power consumption data. The result will be a model that can be used to shortlist consumers who are potentially stealing power, in real time.**

***Keywords*—Non-Technical Losses, Power Theft, Advanced Metering Infrastructure, Neural Networks, Recurrent Neural Networks, Long Short Term Memory, Deep Learning, Machine Learning.**

I. Introduction

The electric power grid is responsible for providing power to more than 80% of the worlds population [1]. However, most grids are based on traditional designs that cannot sustain the growing power requirements of todays society [2]. In India, the traditional power grid is based on designs that are more than 70 years old [3]. An aging power grid is prone to frequent failures as was evident from the major Indian blackout of 2012, which disrupted power supply to more than 60 million people [4]. An unreliable power grid will degrade the life of people dramatically and cease the normal functioning of society.

A potent solution to this problem is the smart grid. A smart grid has many modern features like security, real time demand response, self-healing and two-way transmission [5]. This makes the power grid more reliable. By allowing communication between power utilities and consumers, loads can be fine-tuned and generation overloads and line congestions can be avoided [6]. One way to implement smart grids is to use Advanced Metering Infrastructure. Advanced Metering Infrastructures(AMI's) are an integrated system of smart meters, communication networks, and data management systems that enable two-way communication between utilities and consumers. While research in large scale deployment of smart meters and smart grids is extensive, one aspect that has not been taken into consideration is the detection and prevention of electricity power theft using AMI.

Power Theft is a major issue that is being tackled by Power Utilities today. The World Bank has reported that almost 50% of the generated electricity in developing countries is lost by power theft [7]. These power losses can be classified into two types: Technical and Non-Technical losses. The technical losses are the transmission and distribution losses that can be attributed to defective or aging equipment, internal electrical resistances and malfunctioning components.

The Non-Technical losses include: unauthorised tapping of distribution lines and poles, refusing to pay bills, meter tampering and bypassing meters, bribing of officials and faulty meters [8]. Non-technical losses are inadequately studied and most power utilities do not record data regarding these losses. It is assumed that NTLs are more in developing countries, although even developed countries like USA and UK report NTLs ranging from $1-$6 Billion [9], [10]. In India, Non-Technical losses estimate up to more than 50% of total Power loss. Power utility companies in India have reported a loss of $4.5 Billion incurred annually [11].

The existing technologies for detecting power theft using AMI can be broadly divided into: Artificial Intelligence based, State based and Game Theory Based. Game Theoretic Models is the use of incentive based problems in theft detection when customers cannot be perfectly observed by the distributor [12]. Customer data is kept private as the utility does not know the consumer's power usage perfectly. Game theoretic models are an optimal solution to reduce electricity losses, however, formulating the utility, and customer functions and strategies is a problem.

State based models work by monitoring the state of the power flow to detect power theft. State based detection tech-

978-1-5386-3917-7/17/$31.00

niques were among the first to be implemented to detect theft. Using Wireless Sensor Nodes [13], AMI [8] and RFID tags [14], the power usage can be read. However, most state based models have low efficiencies and require access to customer energy consumption data. This raises the question of privacy. Moreover, the rate of false positives is too high to be properly implemented as a theft detection scheme.

Artificial Intelligence based models are the most widely researched technique to detect NTLs. Power usage profiling of a customer or a group of customers using Support Vector Machines(SVM) [15], Neural Networks [16] and other clustering algorithms is a popular method of detecting theft. In this paper, using the load characteristics of an individual household or a locality, a load profile is created using a Recurrent Neural Network (RNN). This has the potential to reduce the rate of false positives that occur as compared to using traditional machine learning algorithms. If a household uses more or less power than their usual consumption over a period of time, then it is classified as an anomaly.

The paper has been divided as follows. Section II details some common methods of stealing power. In Section III, the intuition behind the theft detection technique is explained. Section IV and V details the architectures used in this paper. Section VI describes the dataset and the data preprocessing techniques used. Section VII explains the details of the proposed model. In Section VIII, a conclusion is drawn.

II. Methods of stealing

The most common method of stealing electricity is by tapping of overhead transmission lines. The theft that has been addressed in this paper is the theft of electric power due to infidelity by the consumer. One common method employed to steal power is by directly connecting the feeder line to the load, or bypassing the Energy Meter completely. Generally, the secondary feeder lines from a meter are extensively insulated. In some cases, the fraudulent consumers remove this insulation and directly connect their unregistered loads and steal power [18].

When the use of Electromagnetic Meters were extensive, consumers would tamper the electromagnetic disc by placing magnets, or a viscous fluid to reduce the rotation of the disc, or stop the disc completely, thus preventing the meter from taking any readings [19]. In countries like India, traditional meters are employed, which face a constant threat of being hacked. Potential hackers interrupt the metering system and tamper with their power consumption data [20].

It is easier to tamper with three phase and two watt hour meters. In two watt hour meter arrangement, by changing the terminals into ground, or removing the ground terminal, or with three phase energy meters the neutral is cut from the distribution feeder, causing the assumption of total energy to be zero.

Apart from this, the type of stealing which is the most difficult to detect, is the illegal consumption of electricity during the peak hours of the day. Hence, smaller household loads are connected with larger loads on the illegal side. Such data is very irregular in nature and very difficult to detect and measure. In some scenarios, employees also meddle with the billing process, showing consumption much lower than the actual.

III. System Model and User Classification

In this section, the intuition of the electricity theft detection system is explained and the description of the system model being considered is given.

It is assumed here, that the electricity system is centralised. Power is generated and transmitted to different loads in a tree-like distribution network. This system has been considered as it is more prevalent in under-developed and developing nations, where the use of distributed energy sources is less. Furthermore, theft in such nations are more as compared to developed countries.

A consumer's load profile or power consumption profile tends to remain the same over a period of time. This is because, large changes in power usages over short periods like weeks or months does not take place. It has been seen that a person's load does not change by more than 5% of the average value [21]. Large changes in power usage, either more or less is associated with power theft.

Power usages greater than the average power can be attributed to line tapping after an electricity meter or after a node in the distribution system. Whereas power usages less than the average value of power can be attributed to a consumer bypassing the energy meter. This results in less power being measured by the utility.

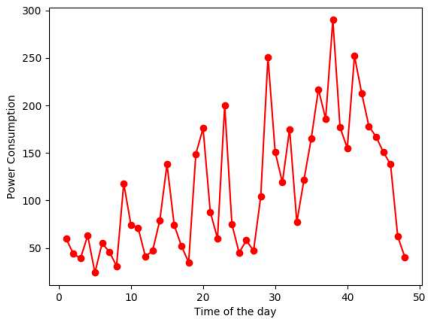

Fig. 1: Load Characteristic of a Power Thief

For a typical user, the consumption of power is very low during the early morning hours. There is a sudden increase in usage during the morning hours which gradually decreases and remains almost constant till the evening. There is a further increase in usage during the dusk-night hours where the power usage is similar to the peak morning hours. This further decreases until midnight.

Compared to a legal user, an illegal user has an erratic power consumption pattern. There is a gradual increase in power in the early hours and an overshoot at the onset of the day. The power usage fluctuates improperly throughout the day.

The peak overshoot is present till midnight when the power gradually decreases.

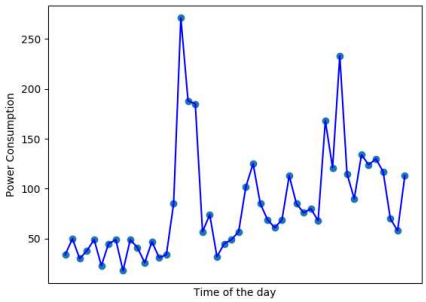

Fig. 2: Load Characteristic of a Normal User

IV. Recurrent Neural Networks

RNNs are a class of Neural Networks that deals with sequential data [22]. Given a sequence of data, it holds memory of all previous computations and uses that memory for future calculations. RNNs, unlike Neural Networks, encode dependency between all inputs. Since RNNs can store memory, they have been very useful in Natural Language Processing and in Speech Recognition.

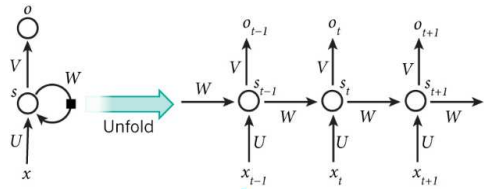

Fig. 3: Framework of a Vanilla Recurrent Neural Network

The above diagram shows how a RNN can be unrolled based on time steps or number of sequences. For example, if the number of sequences is 10, then the RNN can be unrolled into a 10 layer neural network.

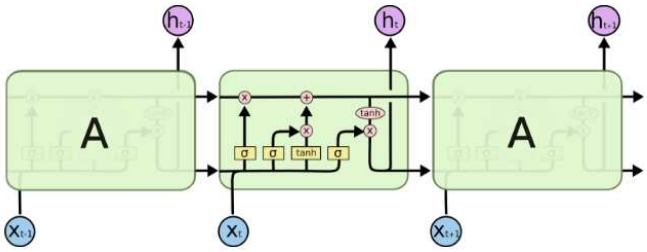

Fig. 4: Long Short Term Memory Architecture [28]

- x_t is the input at time step t. x is a vector of 6 data points which corresponds to the power consumption values for every half an hour.
- U is the weight associated with the input x. This indicates how strong the connection x is with the neuron s.
- s_t is the hidden state which is calculated at time step t and is responsible for the memory. Connection from the previous time step and the current time step are considered for calculating the hidden time step value.
- W is the weight associated with the incoming connection from the previous hidden state.
- The linear combination of $[W \times s_{t-1}]$ and $[U \times x_t]$ is passed through an activation function to calculate the value of the present hidden state s_t. This function is calculated for all hidden states.
- o_t is the output that is calculated by evaluating the *softmax* of s_t. In this case, o_t will be a vector of probabilities, the highest value belonging to the class in which the predicted power consumption value lies.

Training a RNN requires the classic Backpropagation algorithm. This algorithm calculates the derivative of the loss function with respect to weights. These derivatives are then used to modify the weights so as to minimise the loss function. The same algorithm is used to train a RNN with a small variation. In a RNN, the weights are shared across many time steps in the network. Finding the derivative of the loss function with the immediate weight is a naive approach. Hence, the derivative of the loss function is not only calculated with respect to the current time step but also with respect to the previous time steps. If there are 6 time steps and the gradient for the 6th time step is to be calculated, then the gradient of the previous 5 time steps are calculated and added to evaluate the final gradient. Since the gradients are calculated over previous time steps, this method is called Backpropagation Through Time (BPTT).

V. Long Short Term Memory

Long Short Term Memory (LSTM) is a specialised Recurrent Neural Network which has the capability of learning long term dependencies [17] [23]. Traditional RNNs have the problem of vanishing and exploding gradients. In vanishing gradients, weights in the early layers of the network get updated by very low values and hence, train very slowly. This is because weight-updates are proportional to the gradient of the error function and in the initial layers, during Backpropagation, gradients multiply with small activation values repeatedly which further leads to even more smaller values. Similarly, in exploding gradients, the gradients become too large and update the weight parameters by a large value and hence become harder to converge to an optimum value.

$$i = \sigma(x_t U^i + s_{t-1} W^i) \tag{1}$$

$$f = \sigma(x_t U^f + s_{t-1} W^f) \tag{2}$$

$$o = \sigma(x_t U^o + s_{t-1} W^o) \tag{3}$$

$$g = \tanh(x_t U^g + s_{t-1} W^g) \tag{4}$$

$$c_t = c_{t-1} \circ f + g \circ i \tag{5}$$

$$s_t = \tanh(c_t) \circ o \tag{6}$$

TABLE I: Comparison of accuracy on different parameters of architecture

Architecture	Optimiser	Learning Rate	Train Accuracy	Test Accuracy
Single Household				
L(64), D(0.4), L(64), D(0.4), F(6), S	Adam	0.001	73.45%	72.93%
L(32), D(0.4), L(32), D(0.4), F(6), S	Adam	0.001	71.92%	72.5%
L(128), D(0.4), L(128), D(0.4), F(6), S	Adam	0.001	79.51%	70.74%
Locality				
L(10), D(0.6), L(10), D(0.6), F(8), S	Adam	0.001	67.13%	65.3%
L(64), D(0.4), L(64), D(0.4), F(8), S	Adam	0.01	65.55%	65.21%
L(128), D(0.4), F(8), S	Adam	0.001	65.6%	64.73%
L(256), D(0.4), F(8), S	Adam	0.001	67.32%	64.47%
L(256), D(0.5), L(256), D(0.5), F(8), S	Adam	0.001	74.04%	62.66%

LSTMs do not suffer from vanishing or exploding gradients. They have a forget, input and output gate to protect and control the cell state; select what is forgotten, retained in memory and passed as the output. The i (input gate), f (forget gate), o (output gate) and g (candidate) equations are similar. They multiply the input with a weight matrix(U) and the previous time sequences output by another weight matrix(W), sum them and apply an activation function such as *sigmoid* or *tanh* to push the values between a certain interval (0 to 1 in *sigmoid*, -1 to 1 in *tanh*) [24]. The candidate equation, multiplied with the input gate equation, decides how much of input is to be learnt by the cell state and the previous cell states equation, multiplied with the forget gate equation, decides how much of the previous cell state is to be forgotten. The LSTM filters what it outputs from its cell state by multiplying its activated values with the output gate equation.

VI. Dataset

The data used has been collected from an open-source database of electricity consumption benchmarks belonging to the Australian Governments Department of Industry, Innovation and Science [25]. The data consists of power consumption values expressed in kWh for every half an hour of 25 houses for an entire day over a period of approximately 2 years.

A. Sequential Dataset

In this learning model, power consumption values are treated as a set of sequential patterns. The mathematical model, Sequence to Sequence Learning, is used to train the data [26]. As a result of being sequential, the power being consumed in the present depends upon the power that was consumed previously. The same applies to future power consumption values.

B. Intuition Behind Learning

The idea is to use a mathematical model that can learn these patterns and hold some memory of how the data is sequenced. Once trained, the model will take in a fixed vector of inputs and will predict the best possible consecutive value after the sequence. After predicting the output, the window of sequences is slided by a stride of one and the next corresponding power output is predicted. Once all the predicted values are obtained, they are compared with real-time values to find the anomalies.

Initially, the model was trained on data that contained irregularities. Table 1 shows the training results where:

- L(n) = Layer of LSTM (number of nodes)
- D(x) = Dropout(Value)
- F(n) = Fully Connected Layer (number of nodes)
- S = *softmax*

C. Data Preprocessing

The dataset originally contained many entries consisting majorly of zeros, that is, there was no power usage for several hours of that day, and would prove to be noise for training the model. Since the model is based on ideal power consumption values, zeros are removed and outliers are replaced with the maximum value found by box plot.

1) Box Plot: A box-and-whisker plot is used to remove outliers in the data. All power values present in the database of a particular locality is taken in a single vector and sorted in ascending order. The median value is found out and is used to split the data in two parts. Using the median, the data is split and the corresponding medians are calculated for each part. These values are plotted as lower (Q_2) and upper (Q_3) quartiles respectively. Interquartile range (IQR) is found out by the difference in the quartiles $[Q_3 - Q_2]$.The maximum value is calculated by $[Q_3 + 1.5 \times IQR]$ and minimum value by $[Q_2 - 1.5 \times IQR]$. All power values lying outside this maximum-minimum range are considered as outliers. The power values above the maximum value are replaced with the maximum value itself and all power entries that contained majorly 0 are removed for better optimisation and generalisation of the model.

The median calculated for an assumed locality (Discussed in Section VII) in the dataset is 141; lower and upper quartiles are 63 and 336; inter quartile range is 273; maximum and minimum values are 745.5 and -346.5 respectively. The medium calculated for an extracted household in the dataset is 43; lower and upper quartiles are 37 and 74; inter quartile range is 40; maximum and minimum values are 129.5 and -18.5 respectively.

2) Normalization: After data preprocessing has been performed, the data is normalised by dividing each power value by the maximum power value present in the training data.

$$z_i = \frac{x_i}{x_{max}} \tag{7}$$

This value is the maximum value found by the box plot. The power values get squashed in the range $(0, 1]$.

3) Feature adjustment: After data preprocessing and normalisation, all power values in the dataset are unrolled into a single vector. This vector is split into input-output pairs for the LSTM model. A single input will contain 6 continuous power values and its corresponding output will be the next power value in that sequence. The output is then concatenated with the input time sequence as the T^{th} time and the previous $(T-6)^{th}$ value is removed from the input. It's output would be the next power value in that sequence. A stride of *one* is used to move the window over the vector. This is repeated for the entire dataset to form input-output pairs. Intervals ranging from 0-100, 100-200, 200-300, etc. are created. The output value is converted to a categorical value by mapping it to it's corresponding interval. It is then converted to a one-hot representation which is a vector consisting of 1 in it's interval index and 0 in other interval indexes. The input-output pairs are divided into training set, that will be fed to the Recurrent Neural Network for training; validation set, for choosing the optimum parameters and hyper-parameters; and test set, for determining the accuracy of the model.

VII. Model

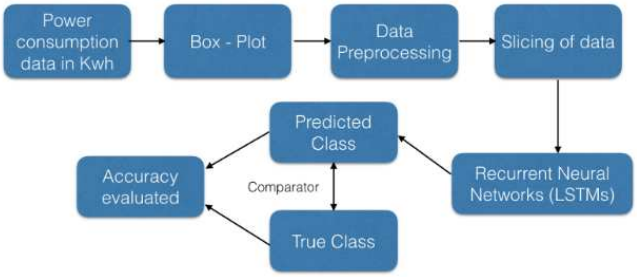

Fig. 5: Proposed Model Architecture

After preprocessing, the data has to be arranged for every locality. The dataset taken provides no information whether the different households belong to the same locality or not. Since the architecture can model on data of a particular locality, we assume 5 households from the dataset that have similar power consumption values as one locality. In the dataset, the rows denote data for a particular house on a particular day and the columns have power consumption values spaced at every half hour intervals. The input dimension is considered as a time sequence of 6 values, thus encoding data about the previous 3 hours.

The proposed model has LSTM layer(s), with each layer having certain number of cells. A dropout layer is added after each LSTM layer. The last LSTM layer is followed by a fully connected layer of neurons which denotes the number of classes. The activation function used in the fully connected layer is *softmax*.

A. Training process

1) Function Evaluation: The model uses Categorical Cross Entropy as the loss function and Adam as the optimiser. Categorical Cross Entropy performs better when compared to Mean Squared Error since Mean Squared Error gives too much emphasis on incorrectly classified examples and may stall training. The final experimental results are optimised with a learning rate of 0.001, fuzz factor of $1e-08$ and no learning rate decay [27].

2) Experimental Values: Fig 6 shows the variation of the loss function with respect to the number of epochs. After 800 epochs, the loss function dropped to 0.5909.

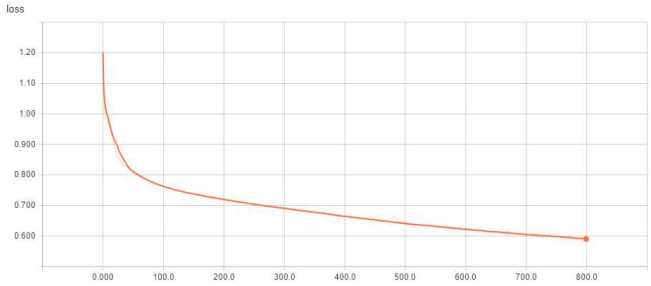

Fig. 6: Training Loss vs Number of Epochs

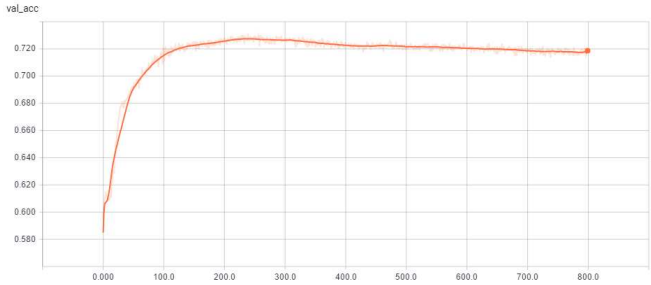

Fig. 7: Validation Accuracy vs Number of Epochs

Fig. 7 shows how the accuracy varies with the number of epochs. The architecture that attained this accuracy consisted of two stacked Long Short Term Memory (LSTM) units consisting of 64 nodes each. A Dropout layer of probability 0.4 is added after each LSTM layer. The last LSTM layer is followed by a fully connected layer consisting of six nodes which denotes the number of classes that the data is being classified into. *Softmax* activation is used in the last layer which scales the output values of the nodes in the range $(0, 1)$. The training accuracy achieved after 800 epochs is 76.81%. Early stopping is used to save the model weights before it starts overfitting. After 227 epochs, the weights are loaded and tested to give 73.45% on training accuracy and 72.93% on test data. The output that the model predicts belongs to classes 1 to 6, where each class denotes a lower bound and an upper bound value. For example, if the output is 2 in one-hot representation, it denotes that the power consumption value should be in the range [101, 200] kWh.

The scalability of the model is arbitrary. Without changing it's architecture, the model can train on not only the data of a whole locality but can also train on a particular household and predict precise values according to the household's power usage profile.

VIII. Conclusion

This paper proposes an Artificial Intelligence based technique of detecting Non-Technical Losses in power grids using Smart Meters. The model detects abnormalities in a consumer's power consumption behaviour and classifies it as an anomaly. The proposed model achieved a test accuracy of 72.93% on a specific household and 65.3% on a locality. However with a cleaner dataset and novel preprocessing techniques, a greater accuracy can be achieved. A cleaner dataset is one without repeating customers, without customers with no consumption (0 kWh) and where the meter readings are recorded regularly and consistently [30]. Considering privacy and confidentiality issues of utilities and customers, proper datasets are not widely available [21]. This is why many approaches have been applied on developed data instead. Machine learning techniques like Decision Trees, K-Nearest Neighbours, Self-Organising Maps and Feed-Forward Neural Networks [29] achieved accuracies as high as 98.4% on such developed datasets [31].

The use of the technique proposed in this paper will help power utilities detect fraud faster and more accurately. This will help utilities handle NTLs better and reduce costs and losses. Further work could be to train the same model on data of a household which has a regular record of stealing power. This would enable the model to learn the patterns of theft and time and hence, predict the hour at which the occurrence of theft is most probable.

References

[1] R. Jiang, R. Lu, C. Lai, J. Luo, and X. Shen, *Robust group key management with revocation and collusion resistance for scada in smart grid, in Proc. IEEE Globe Communication Conference (Globecom)*, 2013, pp. 824-829.

[2] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, *EPPA: An efficient and privacy-preserving aggregation scheme for secure smart grid communications*, IEEE Transactions on Parallel and Distributed Systems, vol. 23, no. 9, pp. 1621- 1631, 2012.

[3] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, EPPA: *An efficient and privacy-preserving aggregation scheme for secure smart grid communications*, *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1621-1631, 2012.

[4] 2012 India blackouts, *http://en.wikipedia.org/wiki/Indiablackout*, 2013.

[5] H. Li, X. Liang, R. Lu, X. Lin, H. Yang, and X. Shen, *EPPDR: An efficient privacy-preserving demand response scheme with adaptive key evolution in smart grid, IEEE Transactions on Parallel and Distributed Systems*, vol. PP, no. 99, pp. 1-10, 2013.

[6] X. Liang, X. Li, R. Lu, X. Lin, and X. Shen, *UDP: Usagebased dynamic pricing with privacy preservation for smart grid*, IEEE *Transactions on Smart Grid*, vol. 4, no. 1, pp. 141-150, 2013.

[7] P. Antmann, *Reducing technical and non-technical losses in the power sector*, Background paper for the WBG Energy Strategy, Tech. Rep., Washington, DC, USA: The World Bank, 2009.

[8] P. Kadurek, J. Blom, J. Cobben, and W. Kling, *Theft detection and smart metering practices and expectations in the Netherlands*, in *Proc. 2010 IEEE/PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*, 2010, pp. 1-6.

[9] P. McDaniel and S. McLaughlin, *Security and privacy challenges in the smart grid*, *IEEE Security & Privacy*, vol. 7, no. 3, pp. 75-77, 2009.

[10] M. S. Alam, E. Kabir, M. M. Rahman and M. A. K. Chowdhury, *Power Sector Reform in Bangladesh: Electricity Distribution System*, Energy, vol. 29, no. 11, pp. 1773-1783, 2004.

[11] Ministry of power, India, Overview of power distribution, Tech. Rep., *http://www.powermin.nic.in, 2013*.

[12] S. Amin, G. A. Schwartz, and H. Tembine, *Incentives and security in electricity distribution networks*, in *Decision and Game Theory for Security*, Springer, 2012, pp. 264-280.

[13] R. V. P. Yerra, A. K. Bharathi, P. Rajalakshmi, and U. Desai, *WSN based power monitoring in smart grids*, in *Proc. IEEE Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, 2011, pp. 401-406.

[14] B. Khoo and Y. Cheng, *Using RFID for anti-theft in a chinese electrical supply company: A cost-benefit analysis*, in *Proc. IEEE Wireless Telecommunications Symposium (WTS)*, 2011, pp. 1-6.

[15] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, *Nontechnical loss detection for metered customers in power utility using support vector machines*, *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 1162-1171, 2010.

[16] C. Muniz, K. Figueiredo, M. Vellasco, G. Chavez, and M. Pacheco, *Irregularity detection on low tension electric installations by neural network ensembles*, in *Proc. IEEE International Joint Conference on Neural Networks*, 2009, pp. 2176-2182.

[17] Hasim Sak, Andrew Senior, Francoise Beaufays, *Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling In Proc. Interspeech*

[18] R. Cespedes, H. Duran, H. Hernandez, and A. Rodriguez, *Assessment of electrical energy losses in the colombian power system, IEEE Trans. on Power Apparatus and Systems*, vol. 102, pp. 35093515, Nov. 1983.

[19] *Pilferage of electricity issues and challenges*, power sector news, KSEB Officers Association, [Online]. Available: http://www.kseboa.org/news/pilferage-of-electricity-issues-and-challenges.html

[20] B. Krebs, *FBI: Smart meter hacks likely to spread, http://krebsonsecurity.com/2012/04/fbi-smart-meter-hackslikely-to-spread/,* 2012.

[21] S. Depuru, L. Wang, and V. Devabhaktuni, *Support vector machine based data classification for detection of electricity theft*, in *Proc. 2011 IEEE/PES Power Systems Conference and Exposition (PSCE)*, 2011, pp. 1-8.

[22] Lipton, Zachary C., Berkowitz, John, and Elkan, Charles. *A critical review of recurrent neural networks for sequence learning. arXiv preprint arXiv:1506.00019*, 2015.

[23] K. Greff, R. K. Srivastava, J. Koutnk, B. R. Steunebrink, and J. Schmidhuber. *LSTM: A search space odyssey. CoRR, abs/1503.04069*, 2015.

[24] Recurrent Neural Network Tutorial, Part 4 Implementing GRU/LSTM RNN with Python AND Theano, *http://www.wildml.com/2015/10/recurrent-neural-network-tutorial-part-4-implementing-a-grulstm-rnn-with-python-and-theano/*

[25] Electricity Consumption Benchmarks, *https://data.gov.au/dataset/electricity-consumption-benchmarks*, Department of Industry, Innovation and Science, Australian Government

[26] Sutskever, I. Vinyals, O. & Le. Q. V. *Sequence to sequence learning with neural networks. In Proc. Advances in Neural Information Processing* Systems 27 31043112 (2014).

[27] W. Zaremba, I. Sutskever, and O. Vinyals. *Recurrent neural network regularization. In arXiv:1409.2329*, 2014.

[28] Understanding LSTMs, *http://colah.github.io/posts/2015-08-Understanding-LSTMs/*, Christopher Olah

[29] P. Glauner, A. Boechat, J. Meira, L. Dolberg, R. State, F. Bettinger, Y. Rangoni and D. Duarte, *The Challenge of Non-Technical Loss Detection using Artificial Intelligence: A Survey, submitted to IEEE Transactions on Power Systems*, arXiv:1606.00626, 2016.

[30] J. Nagi, K. Yap, S. K. Tiong, S. Ahmed, and M. Mohamad, *Nontechnical loss detection for metered customers in power utility using support vector machines, IEEE Trans. Power Del.*, vol. 25, no. 2, pp. 11621171, Apr. 2010.

[31] Depuru, S. S. *"Modeling, detection, and prevention of electricity theft for enhanced performance and security of power grid." (Electronic Thesis or Dissertation). 2012.*