A Hybrid Approach to Power Theft Detection

Saurabh Jain, A. M. Karandikar

Department of CSE, Ramdeobaba College of Engineering and Management, Nagpur, India

Abstract— *Currently power theft is a common problem face by all electricity companies. Since power theft directly affect the profit made by electricity companies, theft detection and prevention of electricity is mandatory. In this paper we proposed a hybrid approach to detect the electricity theft i.e. to detect suspected consumers who is doing theft. We use SVM and ELM for our approach. We also compare our approach with KNN.*

Keywords— ELM, KNN, Power Theft, SVM Classification, Technical Loss, Non-Technical Loss.

I. INTRODUCTION

We all know power theft is a major problem for all electricity companies. This problem is not related to Indian companies only; other country's electricity companies also face this problem. Electricity companies losses money every year due to power theft. There are two types of losses namely transmission loss and non-transmission loss. Transmission loss occurs while transmitting energy form generation side to consumer's side. Following are the some reason for transmission loss occurs:

- Due to improper insulation.
- Due to resistance in wire.

Non-Transmission losses occur due to wrong billing, false meter reading, electricity theft, etc. First two losses can be prevented by taking proper meter reading and calculating accurate bill for electricity consume, but electricity theft is hard to prevent since no one predict about which consumer is honest or dishonest. Still losses due to electricity theft can be kept minimum by finding fraud consumers. There are various ways through which power theft can be done for example bypassing the meter or tempering with meter readings, etc.

Theft detection is done manually by inspecting consumers. This is time consuming process and requires large number of field staff. The cost for this process is too high and detection rate is not so high. To overcome these costs, now a day some data mining techniques are used to detect theft. We are proposing a hybrid approach for detection of theft, which will improve accuracy of detection and requires less cost for whole process.

II. BACKGROUND WORK

Number of methods are proposed and implemented for finding and estimation of power theft. [1] This paper presents a framework to identify power loss activities. They used automatic feature extraction methods for customer profile with ELM, OS-ELM and SVM to identify customer who is doing fraud. They extracted consumption patterns using data mining and statistical techniques. ELM, OS-ELM and SVM classifies profiles for fraud detection. They use outlier detection to find fraud customer profiles, if outlier find and it is due to power loss activity they use this profile as reference. ELM and OS-ELM used as main classifier for their framework. [3] This paper discusses the problems while doing theft detection and previous ways to reduce the theft. In this paper they developed approximate patterns for classification using customer load profiles. Approximate consumption patterns are designed using load profiles and artificial intelligence tools. Then they trained the SVM to classify data based on the suspicious energy consumption. [4] This paper presents a framework to detect non-technical losses. They use Genetic algorithm and support vector machine for their approach. Their approach selects the suspected customers for onsite investigation so theft can be identified.

III. PROPOSED WORK

Machine learning explores the study of algorithms that can learn from and make predictions on data. We proposed hybrid approach to find suspected customers who is doing power theft. We have collected data from IT Office of MSEDCL. This data is a collection of 24 months consumption of customer. Dataset consist fields like consumer number, tariff code, connection load, unit consumption of a month, meter status. We separated some part of dataset as training set and some as test dataset (roughly 80% used for training and 20% used for testing purpose). Our approach contains two main phases namely training phase and data classification phase.

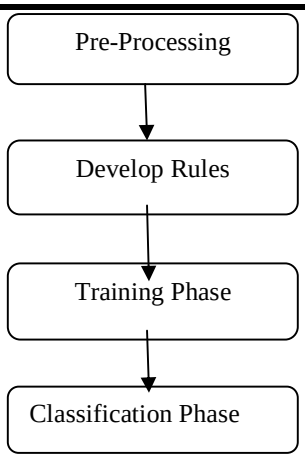

Fig.1: Main steps of the system

1.1 Preprocessing: - In this we classify customers based on tariff code as Residential, Industrial, and Commercial. We selected require fields which is needed for training and remove null values.

1.2 Training: - In this module we analyze data, based on analysis we define rules and develop patterns. E.g. analyze difference between consumptions of successive months, if this difference is greater than some threshold then consider customer as fraud. Using these patterns we train the system. While training we label data as suspected and non-suspected customers. Then we send this labeled data to next phase. We use ELM to find threshold for the training. We calculated threshold using following steps.

1. Considered 10 customers of each type i.e. R, I, C.
2. Calculate difference between average consumptions of two years.
3. Use formula, threshold= ∑ [input * weight] + b.

- **ELM**: It is a feed forward NN with a single layer of hidden nodes, where the weights connecting inputs to hidden node are randomly assign & never updated. ELM not only achieves accurate results but also shortens the training time.

  The ELM algorithm consisting of only three steps that summarized as below,

  1: Assign random weight w and bias.

  2: Calculate the hidden- layer output matrix.

  3: Calculate the output weight.

1.3 Classification: - In this we take labeled data & use defined rules to classify data as non-fraudulent consumer & suspected fraudulent consumer. For classification phase we use SVM. In classification phase we find hyper-plane using labeled data which classifies new data as suspected customers or non-suspected customers. This phase is the last step of the approach.

- **SVM**: A Support Vector Machine (SVM) is a discriminative classifier formally defined by a separating hyper-plane. A good separation is achieved by the hyper-plane that has the largest functional margin.

1.4 Testing:- We use 20% of data for testing purpose. We also implemented k-nearest neighbor for classification purpose so we can compare our approach with KNN. Following table shows the comparison between our approach and k-nearest neighbor for different numbers of customers.

- **KNN:** KNN can be used for both classification and regression predictive problems. However, it is more widely used in classification problems in the industry. In both cases, the input consists of the *k* closest training examples in the feature space. The output depends on whether *k*-NN is used for classification or regression. In k-NN classification, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors (*k* is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of that single nearest neighbor.

Table.1: Comparison between Techniques

Number Of Customers	Time Require For classification (SVM)	Accuracy(SVM)	Time Require For classification (KNN)	Accuracy (KNN)
100	592 millesec	85%	1 sec	90%
200	1 sec	92.03%	3 sec	94.02%
500	4 sec	94.19%	31 sec	95.19%

IV. CONCLUSION

This paper presents a hybrid approach for detection of electricity theft. We use combination of ELM and SVM to detect theft. We collect the data from Maharashtra Govt. electricity utility, based on this data we develop rules and train the system finally we get the list of suspected

customers who is doing fraud as a output. We compare our approach with k-nearest neighbor. Based on comparison we can say that, the proposed approach is fast and also take less time to generate output compare to k-nearest neighbor. As the size of data increases accuracy of our approach also gets increase. This paper also talks about previous work done in this area.

REFERENCES

[1] D.Dangar, S.K.Joshi "Electricity Theft Detection Techniques For Distribution System In Guvnl". International Journal Of Engineering Development And Research | Ijedr (Two Day National Conference (Rteece-2014) -January 2014).

[2] Rong Jiang, Rongxing Lu, Ye Wang, Jun Luo, Changxiang Shen, And Xuemin (Sherman) Shen "Energy-Theft Detection Issues For Advanced Metering Infrastructure In Smart Grid". Tsinghua Science And Technology Issnll1007-0214ll01/12 April-2014 Volume 19, Number 2, Pp105-120.

[3] Soma Shekara Sreenadh Reddy Depuru, Lingfeng Wang, And Vijay Devabhaktuni "Support Vector Machine Based Data Classification For Detection Of Electricity Theft". 2011 Ieee.

[4] J. Nagi, K. S. Yap, S. K. Tiong, Member, *Ieee*, S. K. Ahmed, Member, Ieee, A. Mohammad "Detection Of Abnormalities And Electricity Theft Using Genetic Support Vector Machines". Tencon 2008 - 2008 Ieee Region 10 Conference, Pages 1 – 6, 19-21 Nov. 2008.

[5] Breno C. Costa, Bruno. L. A. Alberto, André M. Portela, W. Maduro, Esdras O. Eler, "Fraud Detection In Electric Power Distribution Networks Using An Ann-Based Knowledge-Discovery Process", International Journal Of Artificial Intelligence & Applications (IJAIA), Vol. 4, No. 6, November 2013.

[6] Paria Jokar, Nasim Arianpoo And Victor C. M. Leung, "Electricity Theft Detection in Ami Using Customers' Consumption Patterns", IEEE Transactions On Smart Grid.