ENHANCING ELECTRICITY THEFT DETECTION ON USER'S METERED DATA

[1]Kalpesh Chaudhari, [2]Mahesh Myadamwar, [3]Manisha Yadav, [4]Shubham Annadate, [5]N.K.Patil
[1]BE Student, [2]BE Student, [3]BE Student, [4]BE Student, [5]Asst. Professor
[1]Department of Computer Engineering, [2]Department of Computer Engineering, [3]Department of Computer Engineering, [4]Department of Computer Engineering, [5]Department of Computer Engineering
Sinhgad Institute of Technology, Lonavala, Pune, India

Abstract— The overall economic growth of India is mostly affected because of endemic electric energy and higher shortages of Electric power supply. Data mining has been used in numerous areas, which include both private as well as public sectors. This paper presents new ways to identify electricity theft by using some intelligence based techniques. The different techniques available for detecting electricity theft are Mobile Remote Checker, Wavelet based feature extraction, Support Vector Machine and Fuzzy based classification techniques. The most of the above mentioned methodologies have performance issues concerning space and time. This paper concentrates on analyzing techniques and to find a proper way to improve Electricity Theft Detection.

Index Terms— Electricity theft, Pre-processing, K-means, Gaussian Distribution, Atkinson Index, ANN and Fuzzy Logic.

INTRODUCTION

Electricity fraud can be defined as stealing of electric power by the means of illegal consumption activities. So, in order to reduce the losses due to electricity theft we need to find the fraudulent customers. But, it is very difficult to identify honest and the fraudulent customers.

Due to increasing population, Electricity distribution system is facing various challenges due excess use of available electricity illegally. This includes electricity theft, meter tampering, unpaid bills, billing irregularities. To find such customers with dishonesty by using efficient measurement has been an active research. This paper presents a framework to identify electricity theft activity i.e. excess consumption of electricity, unpaid bills and faulty meters etc. To identify such dishonest and fraud customer OS-ELM and SVM are used with Automatic feature extraction method. This study uses various user's metered data. Various patterns of electricity used by customer are extracted by using data mining and statistical techniques. Thus we are able to detect fraud activities.

Losses in developing and developed countries can be significantly represented using Nontechnical losses. There are many statistical methods that have been used to check the behavior of the customer in a way that allows detection of NTL activities. The most widely used method is load profiling which is a pattern of electricity demand of the customer. Classification techniques like supervised learning are used in order to extract customers load profile and other data which is useful for NTL analysis. There are many online applications which are needed to be handled, thus OS-ELM has been proposed. The OS-ELM can be used for the real time applications in industries in order to accommodate the data.

For the classification of the given samples we need the Kernel functions present in Support Vector Machine (SVM) that are needed to analyze the similarities between the input samples. Some standard kernels like Gaussian and Polynomial kernels are used to analyze the input samples, but in addition to that there are some kernels which are specially designed to consider structure in the samples. In this paper probabilistic mixture density models are used in order to capture the pattern in data, like Gaussian mixtures for real-valued spaces.

LITERATURE SURVEY

[3] Proposes a method by J.Blom and W.L.Kling that provides insight into the illegal use of abstraction of electricity. The importance and the economic aspects of theft detection are presented and current practices and experiences are discussed. A novel methodology for automated detection by illegal utilization of electricity in the future distribution networks equipped with smart metering infrastructure is proposed. This method may not be much efficient as this proposed system is found to be little bit complex as far as distribution network is concerned. This methodology may save time as helps to maximize profit margin but is complex in its application. In future, this project can be implemented in remote areas, for three phase electric distribution system in India.

[2] Introduces a concept of Smart grid that is being used in order to modernize the electricity distribution system in most of the developing countries. The role of the Smart grid is to monitor, protect and optimize the operations of the interconnected elements automatically. The disadvantage of this system is that it can be easily hacked or hooked, so it is not completely efficient and the cost for the implementation is also very high. Veenetha Pruchuri proposed a mechanism in order to detect the electricity theft by using the BOUNCE algorithm along with some IEEE protocols like p1675 protocol. But this mechanism was only applicable to distribution lines which are highly equipped so the mechanism was not useful for the rural areas. Yang Xioa implemented a technique that discovers the meters which are problematic and provides inaccurate readings. This method was not that efficient because it considers only the nearby meter readings and do not provide the exact location of the theft.

[4] Describes a method that presents the Link method facility and Remote terminal facility in the control room. In the Link method they tried to link main energy meter in the substation transformer and the user energy meter, and proposed that output of user's single phase electric energy meter has an proportional relationship with power. It is not suitable in all cases in all the areas. All this method is based on an above mentioned assumption which is impractical in real case scenarios of electricity distribution and consumption. Also it is very expensive and is not practically possible to reach each and every rural area of the country.

[5] Narrates a method proposed by Christopher and Pravin Thangaraj which is based on Power Line Communication Concept. In the distribution network, High frequency changes its amplitude and frequency due to load in the Electricity Line Changes. Gain Detector is used to detect such a change in high frequency. This can be detected if any activity is done between any two electricity poles and because of that change in Gain occurs and thus we are able to detect theft of electricity and can neutralize that theft but not able to detect improper usage by people and also one main disadvantage for this technique is that it is based on Infrastructure.

[6] Elaborates a method where approach history of user's data usage is used which tends to the creation of customer's information about load profile. Pre-processing of data used by customer is done which includes Data Selection, Data Separation and Data Normalization. After this Important data is automatically taken and then the data is arranged according to abnormal usage of data through the use of ELM and SVM. But SVM is not much accurate so just because of this use of SVM the accuracy of theft detection may reduce.

PROPOSED SYUSTEM

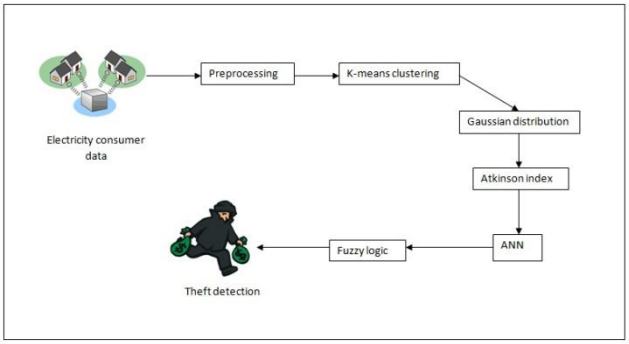

Figure 1: System Overview

Proposed System Implements Five Phase Algorithmic Procedure to predict Theft in Electricity Usage based on Pattern of electricity usage.
Phase 1: Data Reading and Pre-processing: Initially complete data is been read from dataset . Parallel process run to on Training dataset and input Data Set . Only relevant parameters are been selected from input dataset.
Phase 2: K-Means Clustering : This Phase K-means Clustering is been applied on Both Training dataset and input data set generating clusters for furture processing. In K-means Centroid are been Initialized for clustering and based on Euclidean distance measure Data points have been clustered. Outlined Data is also been clustered .
Phase 3: **Gaussian Distribution:** the normal (or Gaussian) distribution is a very common continuous probability distribution. Normal distributions are important in statistics and are often used in the natural and social sciences to represent real-valued random variables whose distributions are not known. In System Gaussian values are been computed to find highly distributed data, Examination has been done on both trained and input data.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{.....(1)}$$

Phase 4: Atkinson Index: Data Inequality has been evaluated for Data measure is useful in determining which end of the distribution contributed most to the observed inequality
Phase 5: ANN and fuzzy Logic: computational model has been designed which is based on a large collection of simple neural units (artificial neurons). This Neurons Assist in Decision Support and communicate with each for Arriving at common decision. ANN Assist in identifying INPUT ,OUTPUT and HIDDEN data in Given dataset. Here in electricity Theft Detection Consumer Data are been sent to neurons for examination. Fuzzy logic Divides Input Data in Five ranges for predictive Analysis here in electricity theft detection Neuron Analysis is been sent to Fuzzy Logic . based On fuzzy Examination Consumers have higher range are been clustered and mostly probably are one who have done theft.

ALGORITHM for Electricity Theft detection

Input: All the attributes
Output: Weight
Step 0: Start
Step 1: Read all the attributes
Step 2: Find n clusters C1, C2, C3, and C4 ….Cn using K means
Step 3: Find SD, mean and Gaussian function of each cluster
Step 4: find minimum range and maximum range of each clusters
Step 5: minimum range =mean
If (Gaussian value > (mean *2))
Maximum range=mean + SD
Else
Maximum range =mean + Gaussian function
Step 6: Apply ANN on C1, C2, C3, and C4…. Cn to generate more clusters by using minimum range and maximum range
Step 7: store all the newly generated clusters to N_C
Step 8: Set Fuzzy parameters
Step 9: For i=0 to N (where N is length of N_c)
Step 10: for each N_{ci} check for d1,d2,d3,d4
Step 11: If (more than 2 attributes are matched)
Step 12: Count++;
Step 13: end for
Step 14: find weight
Step 15: Find Inequality by Atkinson index
Step 16: weight=count/N_c size
Step 17: Stop

IV RESULTS AND DISCUSSIONS

Proposed system of Theft Detection System is deployed as a standalone system using Netbeans as development IDE for java technology which is supported by java media file framework. Performance is evaluated based on the precision and recall parameters. Precision is defined as the ratio of number of relevant Theft detected to the total number of relevant and irrelevant relevant Theft detected. Relative effectiveness of the system is well expressed by using precision parameters. Whereas the recall can be defined as the ratio of number of relevant Theft detected to the total number of relevant Theft are detected not detected. Absolute accuracy of the system is well narrated by using recall parameters.

System can be evaluated using precision and recall parameters, and they can be more clearly elaborated as follows

- X = The numbers of relevant Theft been are detected,
- Y = the number of relevant Theft not detected, and
- Z = the number of irrelevant Theft that are been detected.

So, Precision = (X/ (X+ Z))*100

And Recall = (X/ (X+ Y))*100

Fig.2. Average precision for Theft Detection System

In Fig. 2, by observing it is clear that the average precision obtained for Theft Detection System using ANN is approximately 68%.

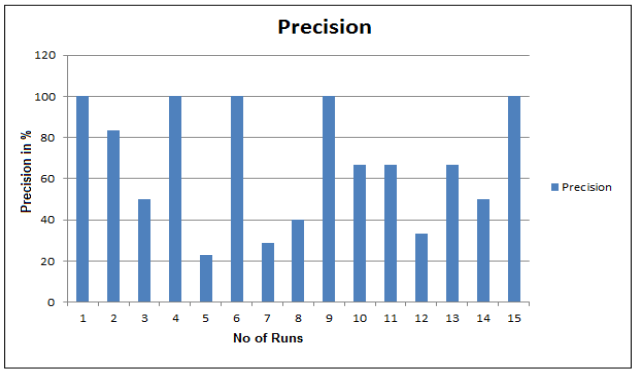

Fig.2. Average precision for Theft Detection System

Figure 3 shows that the system gives 89.33% recall for the Theft detection technique using ANN and Gaussian Distribution. By comparing these two graphs we can conclude that the Theft Detection system using ANN and Gaussian mechanism method gives high recall value compare to the precision value.

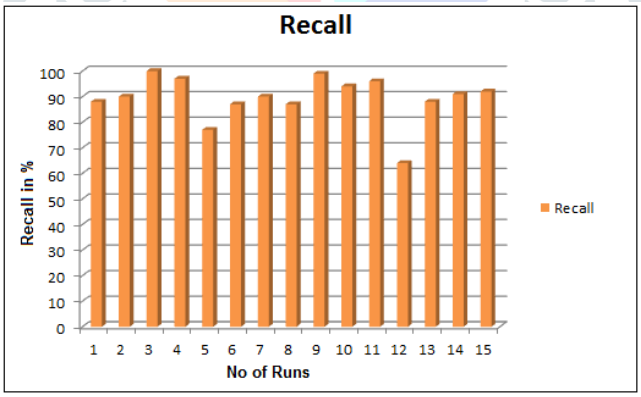

Fig.3. Average Recall for Theft Detection System

V . CONCLUSION AND FUTURESCOPE

Presented Article is Implementation Work on electricity theft detection. Article presents innovation techniques in data mining domain. Presents predictive System which would assist electricity boards to likely detect culprits in Theft . This Research work is purely predictive and hence require confidence and support analysis . furture system can be enhanced and taken to higher level as product . Computational analysis and practical values have been found to be correct predictive and hence system has large scope of research.

REFERENCES

[1] Bharat Dangar, S. K. Joshi, "Electricity Theft detection techniques for Metered Power Consumer in GUVNL, GUJARAT, INDIA", 978-1-4799-1951-2/15/$31.00 ©2015 IEEE.

[2] Selvapriya. C, "Competent Approach For Inspecting Electricity Theft", IJIRSET, Volume 3, Special Issue 3, March 2014.

[3] Mr.M.V.N.R.P.kumar, Ashutosh kumar, A.V. Athalekar, P.G. Desai, M.P. Nanaware, "Electrical Power Line Theft Detection", International Journal of Research in Advent Technology, Vol.3, No.5, May 2015 E-ISSN: 2321-9637.

[4] Kalaivani.R, Gowthami.M, Savitha.S, Karthick.N, Mohanvel.S, "GSM Based Electricity Theft Identification in Distribution Systems", International Journal of Engineering Trends and Technology (IJETT) – Volume 8 Number 10- Feb 2014.

[5] Alwin Vinifred Christopher, Guhesh Swaminathan, Maheedar Subramanian, PravinThangaraj, "Distribution Line Monitoring System for the Detection of Power Theft using Power Line Communication", 978-1-4799-4848-2/14/$31.00 ©2014 IEEE.

[6] Pratap Jumale, Avinash Khaire, Harshada Jadhawar, Sneha Awathare, Manisha Mali, "Electricity Theft Detection Technique: Survey", International Journal of Advanced Research in Computer Science and Software Engineering, ISSN: 2277 128X, Volume 6, Issue 2, February 2016.