

A Model for NTL Detection of Electricity Theft

Dr.S.Nagini^{1*}, B.V.Kiranmayee², K.Jhansi Lakshmi Bai³

¹Associate Professor, CSE Department, VNR VJIET

²Associate Professor, CSE Department, VNR VJIET

³Assistant Professor, CSE Department, VNR VJIET

*E-mail:nagini_s@vnrvjiet.in

Abstract

One of the major factor effecting India's economy is because of power consumer dishonesty. Strategies used for carrying out such activity are like illegal connections, meter tampering, billing irregularities and unpaid bills. Many countries get affected with such problem. Developed model main focus is on fraud detection at Non-Technical Losses (NTL) level. Model has been carried out with the limitations on the type of datasets available. We considered a specific spatial location in Hyderabad (Telangana State). Analysis on such data obtained has been carried out through statistical and k-means clustering approaches. Preprocessing is carried out. The proposed model uses customer consumption data on monthly basis in terms of watts. Onsite inspection rate is reduced from the outcome derived from the model. Analysis figures out abnormal consumption behavior of a consumer. Over years a genuine customer watt usage pattern remains the same for almost all years, with this assumption model was built. Fraudulent consumer were detected if abrupt deviations were found on the consumption pattern over years when compared on the monthly basis. The result of the fraud detection model shortlists potential fraud suspects on whom onsite inspection can be carried out and this information facilitates in reducing the Non-Technical Losses incurred by the electricity board. Electricity theft can be reduced by carrying out onsite inspection or by monitoring such consumers (on the list identified through this model). Or otherwise a clear cut strategy can be followed up by restructuring existing power system distribution or by installing tamper-proof meters that paves way for clear accountability.

Keywords: Non-Technical Losses (NTL), k-means clustering, Preprocessing, Dissimilarity Measures..

1. Introduction

The difference between amount of electricity generated and the amount of electricity distributed or sold calculated in the form of watts is treated as Power Loss [1].

Two Types of power losses are:

- Technical : Power Losses incurred through transformers and cables during power distribution .
- Non-Technical : Losses incurred by tampering watt- meter and illegal connections to derive electricity by consumers[2].

Power generation sector in our state is struggling hard to meet the rise in demand and at the same time the distribution sector is incurring huge amount of losses due to fraudulent consumers[2].

A genuine power consumer is burdened due to losses incurred in this form by government, because government revises power tariff. Based on the survey, conducted it was observed that the city hyderabad has undertaken more number of suspicious electricity robberies in the state. In each and every year the survey represents that more than 250 crores of Electricity is being used illegally by the people.[3] elaborates on various problems in handling these sort of consumers. During the financial 2014-15 power theft (Non-Technical Losses) is said to be around 1.7 per cent[4].

In the recent past it was observed that nearly \$6 billion of electricity is used illegally in the county U.S ,which represents this as one of the most fraudulent activities like the fraudulent activities of credit card data , automobiles and other fraudulent activities [6]. Electricity robbery may cause a huge damage to a country like India [6]. Around 80% of total electricity robbery identified all over the world is from dwelling places and around 20% from

commercial and industrial properties[5]. This paper aims at presenting a model in tackling damages which occurred during systematic application of data preprocessing techniques-collection, reduction, filtering, k-means clustering and also statistical methods. The key factor to progress this research is to support the state electricity boards in decreasing the damage in the area of power distribution sector.

2. Literature survey

The main work carried out in 'Electricity Theft' is in Malaysia to assist Tenaga Nasional Berhad (TNB)[7]. They used data mining approach known as Support Vector Machines (SVMs). There are various machine learning approaches which are utilized for the purpose of identifying the electricity losses. One of the classification algorithm SVM is utilized for classify the customers as fraud and non-fraudulent based on the load profile[14,15]. Whenever there is abrupt changes in the load profile, such sort of consumers are identified as fraud. Load profiles of the customers is taken into account are for every 15 minutes of the day. So 2,880 readings are available for single customer[7]. This feature can be implemented if load profiles can be registered perfectly between distribution and utilizing points. But for the data collected for creating this model such load profile perfect maintenance is missing , thereby lot of missing values are sustained.

The frame work projected for the sake of NTL analysis uses the classification algorithm SVM in identifying the customer usage data of consumption[19]. From the projected approach we can identify the illegal usage of power consumption by the people.

This frame work projected[5] in recognition of losses occurred in the county malaysia.In identifying the losses they have used Remote meters which they can identify data of power consumption data using remote meter events wirelessly to TNB Metering Services. The Remote Meter Reading (RMR) it record data of daily consumption data .This approach finds the losses bySVM and Fuzzy Inference System.

3. Existing system

The current approach followed by Electricity board to reduce non-technical losses is they perform onsite inspection of consumers. Problems with existing system:

- It is a challenging task to identify untrue consumers.
- On-site inspection of fraudulent activities involves enormous operational costs.
- Lack of grid discipline(load profile calculation) So, need for an automated system for fraudulent consumers detection is immense.

4. Proposed System

The proposed model provides an automated intelligent system in identifying the losses occurred in power consumption and distinguishing the untrue consumers depending on design of watt consumption [15] of customers which exists in the consumer database. This proposed method is useful in identifying the untrue transactionsThis unique method is proposed in recognizingthe fraudulent customers and the illegal patterns of power consumption behavior.

In recognizing the data of usage of customers we are using various Statistical Methods. The usage of consumers interesting designs are mined through various data mining methods, which characterizes the consumer load summaries information. Depending on the hypothesis load summaries informationirregularities can be identified when a fraud event taken place. The irregular consumption scenario of the customer is identified by the applied methodology. There are numeroustypes of fraudulent activities can be taken place but we focusses only on situations where sudden changes seem to be occurred in the consumers behaviour.Proposed system assumes that the pattern of Watt usage w.r.t to each customer is same for every year. To identify the relative difference of year's watt usage of every customer, dissimilarity measure [11,20] is calculated for every customer with the relation of "More the measured Dissimilarity more probable to become Fraud".

5. Modules related to the model

Belowgivensteps are automatedbyusing a server and a backgroundexecution of allthe modules, therebygettingtheresults at theeaseon a single click.

The module automationisimplemmentedusingTomcat as a test server and Java Programming to implementallthe modules.

- Database Module: as shown in Fig 1

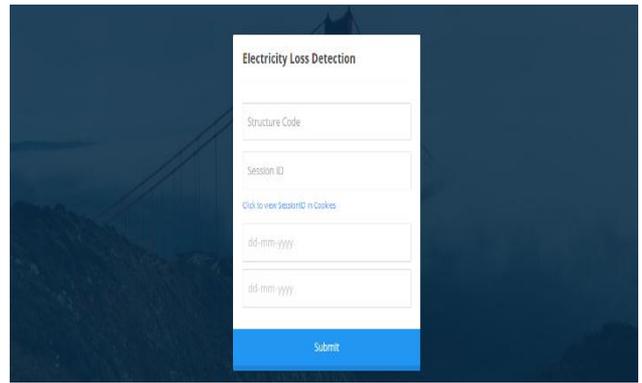


Fig 1: Database Module

- Data Acquisition: as shown in Fig 2

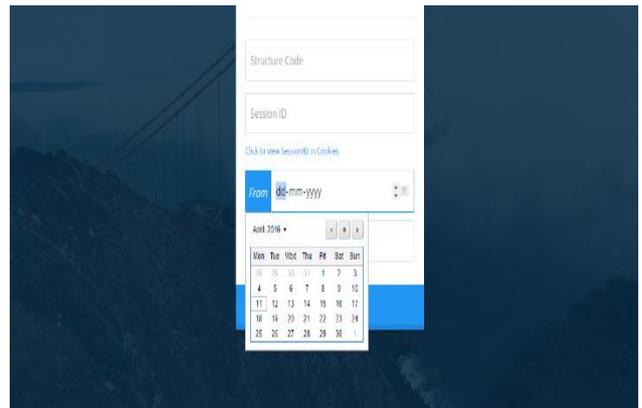


Fig 2: Data Acquisition

- Customer Filtering and Selection as shown in Fig 3
- Feature Extractionas shown in Fig 3

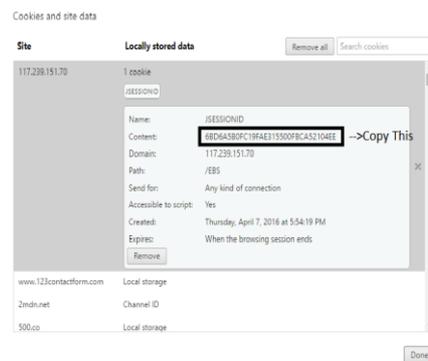


Fig 3

- Data collection from website as shown in fig 4
- Data collection from website is implemented by using cookies.

Adding theft informationwebsite as shown in fig 4

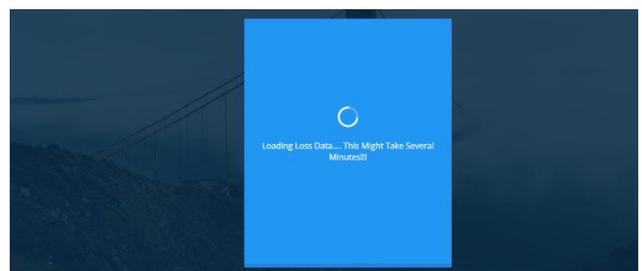


Fig 4

After adding theft information, the statistical methods are applied on the data using MATLAB.

The Matlab execution is also automated using Java Matlab Control API.

Statistical Methods

- Year DisSimilarity
- Normal Probability Plot
- Social Choice Theory

Grubb's Test for Outlier

Clustering Methods

- k-Means

Dissimilarity procedures are one of the potential methods for identifying the features which can be used to calculate the distance between the objects and we can define the method based on the quality of the feature performance[20].

Year Dissimilarity Metric:

It is expected using the pattern of Watt usage w.r.t to each customer is same for every year. To identify the relative difference of year's watt usage of every customer, dissimilarity measure is calculated for every customer with the relation of "More the measured Dissimilarity more probable to become Fraud".

Formula:

Total Dissimilarity = Dissimilarity(Y1,Y2 + Y1,Y3 + Y1,Y4 + Y2,Y3 + Y2,Y4 + Y3,Y4)

Experimental Results:

Top 99 includes identified Frauds. (47% records Eliminated to check)

NPP (Normal Probability Plot)Process

The NPP process is generally is a graphical method which is used to recognize a data set whether it is normally distributed or not [21]. Here, sorted data vs. Expected values are plotted which shows that the image appearances like a straight line as shown in Fig 5.

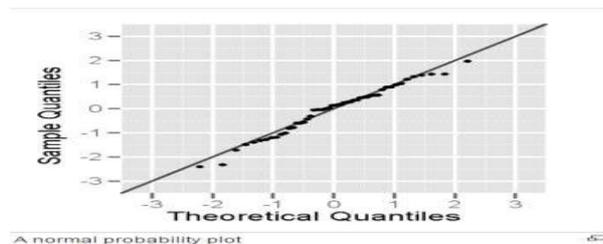


Fig 5

If the information are generally scattered. Deviations from a straight line recommends that it withdraws from normality [21,22]. [For Experiment $E[X] = \text{mean value.}$]

Experiment:

1. Project points on line $y=x$
2. Calculate the Dispersion of points that satisfies (Actual-Expected < 0 i.e., customers who uses less value than expected).
3. Sort according to Dispersion values.

Social Choice Theory (Voting) Application for data approximated as Normal Distribution:

Assumption: Every Month watts Consumption by customer follows normal Distribution.

A Normal distribution in a variate X with is a static distribution with probability density function.

Experiment:

1. Plot Normal Distribution Curve for every month and calculate $p(x)$ for every customer w.r.t every month using above formula.

Now representing the matrix data:

C1 p1-1(y) p1-2(x) p1-3(x) p1-60(y) classID

C2 p2-1(y) p2-2(x) p2-3(x) p2-60(y) classID

.

.

Cn pn-1(y) pn-2(y) pn-3(y) pn-60(y) classID

Here probability Density function (pdf) value is calculated using formula:

From above formula less the y from mean, more the diff "y-mu" and less $p(y) \Rightarrow$ more fraud assumption.

2. Rank the Customer for every month with less $p(y)$ as less rank. (Now we have Different Customer Rankings for different months.)

3. Use Concept of Social Choice Theory Voting Methods from "Game Theory" to decide the total Ranking of Customer [22].

SCT (Social Choice Theory) Method:

It is a method which reads decision problems in which a group has to make a choice .It generally not representing theory but consisting of group of methods and outcomes regarding the combination of the individual inputs and their preferences.

Eg: $m > n > o = p$ is read as "m is favored to n which is favored to o which is equivalent to p"

Decision Taking methods in this scenario includes:

- ✓ Condorcet Principle
- ✓ Baroda Count

Baroda Total:

- Individually ballot is a list of well-ordered alternatives.
- separately calculate the rank of each alternative.
- Rank order replacements bases on declining sum of their ranks.

In the current Experiment:

Each month is: Voter

Ranks of pdfs of every month: Rank order of alternatives by voter.

Result: Top 104 includes identified Frauds. (44% records Eliminated to check)

Grubb's Test for Outlier

It is a graphical or statistical method which is utilized in identification of outliers. The method generally calculates and finds out outliers in a normal population. Using the Grubbs' test we can identify one outlier at a time. After identifying the outlier it is removed from the dataset and the procedure is iterated until no outliers are perceived.

Experiment:

1. Give the order of Outliers while iterating w.r.t to every month.
2. Identify the total rank of customer becoming outlier w.r.t all months and order accordingly.

Clustering (k-means): An algorithm that organises data into groups which are quiet similar , the remaining objects which fall outside the cluster are called as outliers . The algorithm is as follows[23].

Algorithm :k-means. The k-means algorithm for partitioning, where each cluster's center is represented by the mean value of the objects in the cluster.

Input :

k: the number of clusters,

D: a data set containing n objects.

Output : A set of k clusters.

Method:

1. Arbitrarily choose k objects from D as the initial cluster centers;
2. Repeat
3. (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
4. Update the cluster means, i.e., calculate the mean value of the objects for each cluster;
5. Until no change;

Experiment:

1. Cluster all points for k=2 using k-means
2. In every cluster identify the distance of sample w.r.t mean and sort points according to increased distance w.r.t mean.

The approach is mainly based on the customer load profile on monthly basis. Data is collected from Telangana State Southern Power Distribution Company Limited (TSSPDCL). For each customer five year data has been taken for experiment. So total 60 records for each customer is available.

Experimental Results : -

- 1) Year Dissimilarity Method : 53% of records has to be checked for identification of genuine customer.
- 2) Normal Probability Plot : This method is discouraging because 80% records have to be checked for fraud. In practicality it is not possible to verify . So this method fails to identify the fraud.
- 3) Social Choice Theory : 56% of total available records has to be checked.
- 4) Outlier Detection method : 50% of available records are identified as fraud.
- 5) K-Means Clustering : 60% of records has to be checked.

TL	SERVICE NO.	CAT	NAME	METER NAME	TC	SEAL
18	000001	1	P.O. CHITRAL			
19	000002	1	P.O. CHITRAL			
20	000003	1	P.O. CHITRAL			

Fig 6

Based on the above result (Fig 6) we conclude that any way 50% of the customers have to be inspected. What technique is implemented in the model is elided. The total data and the details of the suspected fraudulent consumers data has been shared with relevant authorities and hence it has not been included in this paper.

6. Conclusion

This paper projected a method that automates statistical and data mining techniques in identifying the losses which are occurred through electricity usage. This projected method for NTL analysis

is used to recognize the irregular performances which will be quite helpful for State Electricity Boards. Different statistical measures, Outlier Detection methods, Clustering approaches are used in identifying the fraudulent activities. Suspected fraudulent identified consumers can be inspected onsite for verification. Tamper-proof meters can be installed both at consumer point and at the distribution point, this gives accountable information leaving apart the minimum percentage of transmission losses.

7. Future Scope

1. By knowing the load profiles as in Malaysia fraud detection can become much more accurate.
2. This model can be extended to various commercial and industrial areas.
3. The model can be extended to other departments such as water, sanitation, telecom etc.
4. The same model can be implemented using multiple class SVM to achieve higher accuracy if the datasets are properly maintained.
5. Through modelling can bring down a good plan for deploying power system which enables in reducing electricity theft.

References

- [1] Oda Refou, QaisAlsafafef, Mohammed Alsoud " Evaluation of Electric Energy Losses in Southern Governorates of Jordan Distribution Electric System", IJEE 2015, 5(2):25-33, DOI:10.5923/j.ijee.20150502.02
- [2] MehebabAlam, Sk Mohammed, Mandela Gain " A Review of Losses in Distribution Sector and Minimisation Techniques", IJAREEIE, ISSN(Print):2320-3765, Vol. 3, Issue 10, October 2014.
- [3] An article " Hyderabad Tops Power Theft List in AP" www.deccanchronicle.com on Jan 5th 2014
- [4] An article "Telangana, Andhra Pradesh discoms' reports under scanner" in www.deccanchronicle.com on Jan 4th 2015
- [5] http://en.wikipedia.org/wiki/Theft_of_electricity
- [6] <http://www.forbes.com/sites/peterdetwiler/2013/04/23/electricity-theft-a-bigger-issue-than-you-think/>
- [7] R. Jiang, H. Tagaris, A. Lachsz, and M. Jeffrey, "Wavelet Based Feature Extraction and Multiple Classifiers for Electricity Fraud Detection" in Proc. of IEEE/PES Transmission and Distribution Conference and Exhibition 2002: Asia Pacific, Vol. 3, pp. 2251-2256.
- [8] C. R. Paul, "System loss in a Metropolitan utility network" IEEE PowerEngineering Journal, pp. 305-307, Sept. 1987.
- [9] I. E. Davidson, A. Odubiyi, M. O. Kachienga, and B. Manhire, "Technical Loss Computation and Economic Dispatch Model in T&D Systems in a Deregulated ESI" IEEE Power Eng. Journal, Apr. 2002.
- [10] www.cea.nic.in/reports/monthly/executive_rep/jan12/1-2.pdf
- [11] J. W. Fourie and J. E. Calmeyer, "A statistical method to minimize electrical energy losses in a local electricity distribution network" in Proc. of the 7th IEEE AFRICON Conference Africa: Technology Innovation, Gaborone, Botswana, Sept. 2004.
- [12] J. R. Filho, E. M. Gontijo, A. C. Delaiba, E. Mazina, J. E. Cabral, and J.O. P. Pinto, "Fraud Identification in Electricity Company Customers Using Decision Trees" in Proc. of 2004 IEEE International Conference on Systems, Man and Cybernetics, Vol. 4, pp. 3730-3734, Oct. 2004.
- [13] J. R. Galvan, A. Elices, A. Munoz, T. Czernichow, and M. A. Sanz-Bobi, "System for Detection of Abnormalities and Fraud in Customer Consumption" in Proc. of the Electric Power Conference, Nov. 1998.
- [14] A. H. Nizar, Z. Y. Dong, and J. H. Zhao, "Load Profiling and Data Mining Techniques in Electricity Deregulated Market" IEEE PowerEngineering Society General Meeting, 2006, 18-22 Jun. 2006, pp. 1-7.
- [15] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "Allocation of the load profiles to consumers using probabilistic neural networks" IEEE Transaction on Power Systems, Vol. 20, No. 2, pp. 548-555, May 2005.

J. H. Holland, Adaptation in Natural and Artificial Systems. Ann

- Arbor,MI: Michigan Univ. Press, 1975 (Cambridge, MA: MIT Press, 1992)
- [16] Y.-K. Kwon, B.-R. Moon and S.-D. Hong, "Critical heat flux function approximation using genetic algorithms" IEEE Transactions on Nuclear Science, Vol. 52, No. 2, pp. 535-545, Apr. 2005.
- V. Vapnik. Statistical Learning Theory, John Wiley & Sons, 1998.
- [17] D. Wang, D. S. Yeung, E. C. C. Tsang, "Weighted Mahalanobis Distance Kernels for Support Vector Machines" IEEE Transactions on Neural Networks, Vol. 18, No. 5, pp. 1453-1462, Sept. 2007.
- [18] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [19] "Statistical Learning with Similarity and Dissimilarity Functions" vorgelegt von Dipl.-Math. Ulrike von Luxburg an der TU Bingen, <https://www.informatik.uni-hamburg.de/ML/contents/people/luxburg/publications/>
- [20] <http://www.itl.nist.gov/div898/handbook/eda/section3/normprpl.htm>
- [21] <http://home.uchicago.edu/rmyerson/research/schch1.pdf>
- [22] <https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm>