Electricity Theft Detection Techniques for Urban Commercial Power Consumers

Bharat Dangar, S. K. Joshi

***Abstract:** In power utilities, including technical losses and those due to Unmetered Power Losses activity in urban commercial consumes, and relating to the impact of Unmetered Power Losses activity from a financial perspective and an economic. Finally, this paper provided an overview of MGVCL as the largest power utility in Gujarat and outlines its needs with respect to implementing solutions to minimise Unmetered Power Losses activity.*

***Keywords:** Mgvcl, Load Profile, Dataminig, Electricity, Technical Loss, Cost, Kwh*

I. INTRODUCTION

In many countries electricity consumption profiles recognized as a substitute, price-signal approach to the mean time metered solution which is inappropriate and costly for lower & medium voltage, domestic & trading consumers. Power consumers load profiles helps power utilities to determine the power cost, it helps Power utilities to improve efficiency, planning & trading approach.

II. LOAD PROFILE AND DATAMINING TECHNIQUES

Clustering techniques, known as unsupervised learning, provide starting procedure for solution in examining data analysis and add up with pattern recognition methods. The general purpose of cluster analysis is to identify frequent patterns or to add up related cases through a process of incorporating a set of objects into clusters. In first cluster object should be similar comparing the object in another cluster should be dissimilar is intended outcome. This procedure is not only useful tool for finding the distribution of patterns and interesting correlations among data attributes, but it also acts as an outlier detection tool to identify and detect objects that deviate from normal patterns [1].

In many real-world applications Clustering techniques have been widely used including document clustering, gene expression micro-array data analysis, and image segmentation. Also, they have commonly been used in power utility applications, particularly in load profiles studies, to group analogous load profiles for various purposes. These have included developing better marketing strategies, properly designing tariff structures, and allocating typical load profiles (TLPs) to form groups of eligible consumers. Up till now, no study uses a classification process to group load patterns for each individual consumer based on behavioral similarity as a means of establishing normal and abnormal load patterns for identification and detection benchmarking purposes.

Revised Manuscript Received October 05, 2019

Bharat Dangar, Electrical Engineering Department, Faculty of Technology and Engineering, The M S University of Baroda

S. K. Joshi, Electrical Engineering Department, Faculty of Technology and Engineering, The M S University of Baroda

III. ELECTRICITY LOSSES

In general, affected electricity utilities of power losses are categories into two categories: nontechnical losses and technical losses. Power losses are defined as the difference between powers supplies quantities recorded as sold to consumers and power supply quantities delivered. The power registered as consumed should equal the electrical power generated. In real scenario, the condition is not similar because losses occur as an integral outcome of distribution losses and energy transmission. Davidson [2] developed these power losses in terms of the as following equations.

Power Losses is defined as follows:

$$E_{Loss} = (E_{Delivered} - E_{Sold})$$

Technical losses arise because of the physical climacteric of power generation, T & D and turbine efficiency covers degrees in generation, transformer, together with substation and line related losses. These involved resistive losses in the primary line (IR), resistive losses and the distribution transformer losses (windings and resistive losses in core losses) in the secondary line, losses in kWh metering and resistive losses in service drops.

Credit loss because of technical losses:

$$C_{Loss} = (U_{Electricity\ Cost} \times E_{Loss}) + M_{Maintenance\ Cost}$$

In general, we can categories in two types of technical losses: i) the no-load losses are independent of the load served by the system. The major no-load losses are because of transformer core losses leads to excitation current flows and ii) load losses consisting of the I2R and I2X losses in the series impedances of the various system elements, although when the system is unloaded, these load losses are obviously non-existent. Unmetered Power losses corresponds to power theft in one form or another which are related to the consumer management process and can include several means of consciously defrauding the utility concerned [3].

Unmetered Power Losses (UPL)

$$C_{Unmetered\ Power\ Loss} = (C_{Loss} - C_{Technical\ Loss})$$

Several of this study used data mining algorithm by directly determine them to consumer databases as inputs. A combination of multiple classifiers and wavelet techniques have been applied to identify fraud in a power distribution network. Maximum accuracy is obtained with wavelet technique over easy methods because of its capacities in multi and localization resolution study. As another option, rough sets & decision tree were used respectively for the classification of power consumers. Study also conduct using statistical based outlier mining and the artificial neural network [4], where both studies review a method employing a common framework that had consumer databases as its input data source.

3.1 Data cleaning

Data cleaning is the solution for missing or incomplete data values and for inconsistent and noisy data. In such cases, it functions by filling in missing values, smoothing noise, and resolving inconsistencies. These issues may result from human mistakes or from equipment faults that can affect the quality of data pre-processing and thus affect the quality of the results of subsequent data mining tasks.

3.2 Data Transformation

Normalization techniques may improve data mining results as they scale the measured values to a specific range, for example [-1, 1] or [0, 1]. In power utility customer data such as that accumulated by MGVCL, some commercial customers may have larger consumption values that can outweigh other smaller-scale consumption patterns of customers. The result is a bias in distance measures. Therefore, the customer load data requires normalization for pattern comparison purposes and this forms one of the data pre-processing stages.

3.3 Data Integration

Data mining tasks most often involve multiple databases and data integration can merge data from these disparate sources. The data is segregated according to the load conditions and divided into groups of similar consumers presenting similar characteristics. It is necessary to classify the data in terms of the days of the week, as different load-shape patterns for working days and weekend days.

3.4 Data selection

MGVCL Gujarat data required for this study consists of the accumulated records of 46 as shown in table 1 commercial customers from 2013 to 2014.Data separation – separation of the customer data is based on types of days of the week from Monday until Sunday with additional citations for public holidays hourly load data.

Table 1: Number of customers provided by MGVCL, Gujarat based on the two-city selected

City	Area	No. of Customer	Mean	Std. Deviation
BARODA O&M CIRCLE	Urban	23	104.10	7.455
	Jyotigram	06	105.39	7.219
ANAND O&M CIRCLE	Urban	13	102.00	7.566
	Jyotigram	04	103.05	5.785

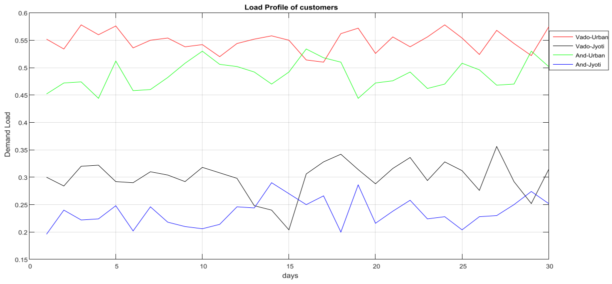

Figure 1 Customers' load profiles

Case-1- If the power consumption is zero from 1 to 30 days, then it is absolutely an unmetered power loss due to faulty metering. Case-2 - If the power consumption within the range generated by the clustering techniques, then the load power consumption is fall in normal behavior on that day. Case-3 - If the power consumption detects outside of the range generated by the clustering techniques, then the load power consumption is fall in suspicious behavior on that day. Its required on-site field investigation and it is generating an alert to maintain department for observation.

Procedure: The following steps make up the classification process. Train the customer data – a set of individual customers' data is used, with the datasets separated according to types of days. Apply the test data – a different set of customers' data separated by types of days is supplied for the testing procedure. Compare the classification accuracies measured as percentages and the time processing durations measured in seconds.

1. OS-ELM

A sequential learning algorithm referred to as online sequential extreme learning machine (Online Sequential-Extreme Learning Machine) [07, 06, 08] that can handle both additive neurons and RBF nodes. It was developed to minimize the limitation of Extreme Learning Machine as proposed by Huang. As the Extreme Learning Machine algorithm belongs among the group learning algorithms, this limited its more application. In a practice, the training data may arrive one-by-one and the online sequential learning is required to provide for such variety. In addition, some online industrial applications prefer sequential learning algorithms as they do not need to retrain whenever a new dataset appears. So far, two types of Online Sequential-Extreme Learning Machine have been proposed: i) Online Sequential-Extreme Learning Machine based on recursive least squares (RLS), and ii) improved Online Sequential - Extreme Learning Machine known as (OLS) for Online Sequential-Extreme Learning Machine (RLS).

Another comprehensive analysis applying the OS-Extreme Learning Machine with the sigmoid function was conducted using the same datasets as used in the Extreme Learning Machine with different numbers of hidden neurons from 20 to 200. The OS-Extreme Learning Machine with sigmoid function results are separated into two tables. Table 2 presents the time processing durations in seconds. Table 3 presents classification success rates as percentages. The best classifier for each type of days is chosen based on the highest classification accuracy ascertained from the simulations.Table 2 shows the results for time processing durations in seconds applying the OS-Extreme Learning Machine with the sigmoid function. Datasets with smaller hidden numbers of neurons produced faster time processing speeds compared to datasets with larger hidden numbers of neurons. The larger the number of hidden neurons, the longer the time processing duration in seconds. Overall, minimum 20 hidden neurons are the most suitable and gave the fastest results for all types of days, 200 hidden neurons = 0.733 seconds and 20 hidden neurons = 0.095 seconds. As show in table 3 have higher classification accuracies with the number of hidden neurons equal to 200. The best number of hidden neurons is found to be 20 with 94.73%

accuracy, it is 200 with 95.98% accuracy, and it is 20 with 94.70% accuracy.

Table 2: Results for time processing durations in seconds with OS-ELM

No. of hidden neurons	Training (secs)	Testing (secs)
20	0.095	0.095
40	0.145	0.132
60	0.158	0.141
80	0.237	0.178
100	0.373	0.261
120	0.463	0.367
140	0.589	0.422
160	0.655	0.494
180	0.687	0.573
200	0.865	0.733

Table 3: Results for classification accuracy as percentages with OS-ELM

No. of hidden neurons	Training (%)	Testing (%)
20	94.74	94.73
40	94.22	94.06
60	94.34	94.14
80	95.69	94.27
S100	95.88	94.96
120	95.93	95.05
140	96.08	95.05
160	96.24	95.36
180	97.26	95.58
200	97.40	95.66

Table 4 Root Mean Squared Error Results With The Prediction Algorithms Based On Different Season

Commercial Customers	OSELM Sigmoid		OSELM Radial basis function	
	Training	Testing	Training	Testing
Urban	0.0283	0.0359	0.0438	0.0455
Jyotigram	0.0274	0.0571	0.0284	0.0280

It is apparent from Table 4 that different prediction algorithms generate different error rates. For Urban and Jyotigram, Extreme Learning Machine with the sigmoid function gives the lowest error rates. With respect to the higher error rates, ELM radial basis function produced the highest error rates for Urban and Agriculture, from these observations, it can be concluded that the sigmoid activation function produced lower error rates when compared to the radial basis function nodes function.

IV. CONCLUSION

Online Sequential - Extreme Learning Machine algorithm in online sequential learning mode, Prediction accuracy is found better for urban commercial customers and it generated better results in terms of the actual classification rates, it will speed up the classification process for urban commercial abnormal behavior customers datasets, Online Sequential-Extreme Learning Machine with the sigmoid function produces lowest error rates in forecasting load results. Online Sequential-Extreme Learning Machine with radial basis function nodes produces the highest error rates in forecasting load results, based on types of seasons.

REFERENCES

1. Shelke, M. Thakare, P.D. Short Term Load Forecasting by Using Data Mining Techniques. Int. J. Sci. Res. 2014, 3, 1363–1367.
2. I. E. Davidson, "Evaluation and effective management of non-technical losses in electricalpower networks," presented at the IEEE AFRICON Conference in Africa, George, South Africa, Oct. 2002.
3. I. E. Davidson, A. Odubiyi, M. O.Kachienga, and B. Manhire, “Technical Loss Computation and Economic Dispatch Model in T&D amp; Systems in a Deregulated ESI” IEEE Power Eng. Journal, Apr. 2002.
4. Inigo Monedero, Felix Biscarri, Carlos Leon, Jesus Biscarri, and Rocío Millan, “Detection of Non-Technical Losses in Electrical Consumption Using Neural Networks and Statistical Techniques”, ICCSA, LNCS 3984, pp.725 - 734, 2006.
5. K. Y. Lee, Y. T. Cha, and J. H. Park, "Short-term load forecasting using an artificial neural network," IEEE Transactions on Power Systems, vol. 7, no. 1, pp. 124-32, Feb. 1992.
6. N.-Y. Liang, G.-B. Huang, P. Saratchandran, and N. Sundararajan, "A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks," IEEE Transactions on Neural Networks, vol. 17, no. 6, pp. 1411-23, Nov. 2006.
7. Nan-Ying Liang, Guang-Bin Huang, P. Saratchandran and N. Sundararajan, “A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks” IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 17, NO. 6, NOVEMBER 2006.
8. B. Li, J. Wang, Y. Li, and Y. Song, An Improved On-Line Sequential Learning Algorithm for Extreme Learning Machine, Berlin: Springer Berlin Heidelberg, 2007, pp. 1087-93.