Electricity Theft Detection Techniques for Distribution System in GUVNL

[1]D.Dangar, [2]S.K.Joshi

Electrical Engineering Department
The Maharaja Sayajirao University of Baroda , Vadodara, Gujarat, India
[1]dangarbharat@yahoo.com, [2]skjoshi@ieee.org

***Abstract*—Electricity consumer dishonesty is a problem faced by all power utilities. Finding efficient measurements for detecting fraudulent electricity consumption has been an active research area in recent years. This paper presents a new approach towards Distribution Power Loss analysis for electric utilities using a novel intelligence-based techniques like Extreme Learning Machine (ELM), OS-ELM & Support Vector Machine (SVM). The main motivation of this study is to assist Gujarat Urjha Vikas Nigam LTD (GUVNL) to reduce its Distribution Power Loss due to electricity theft. The proposed model preselects suspected customers to be inspected onsite for fraud based on irregularities and abnormal consumption behaviour. This approach provides a method of data mining and involves feature extraction from historical customer consumption data. The approach uses customer load profile information to expose abnormal behaviour that is known to be highly correlated with Power Loss activities. The result yields classification classes that are used to shortlist potential fraud suspects for onsite inspection, based on significant behaviour that emerges due to irregularities in consumption. Simulation results prove the proposed method is more effective compared to the current actions taken by GUVNL in order to reduce Power Loss activities.**

***Index Terms*—Support vector machine, intelligent system, electricity theft, Distribution Power Loss, load profile.**

I. INTRODUCTION

ELECTRIC utilities lose large amounts of money each year due to fraud by electricity consumers. Electricity fraud can be defined as a dishonest or illegal use of electricity equipment or service with the intention to avoid billing charge. It is difficult to distinguish between honest and fraudulent customers. Realistically, electric utilities will never be able to eliminate fraud. It is possible, however, to take measures to detect, prevent and reduce fraud [1].Investigations are undertaken by electric utilities to assess the impact of technical losses in generation, transmission and distribution networks [2-3]. Distribution Power Loss comprises one of the most important concerns for electricity Utilities worldwide.

Table 1GUVNL : Overall Distribution Losses in % as on Jan-12

Category	08-09	09-10	10-11	April-10 to Jan-11	April-11 to Jan-12	Modified % as on Jan-12
HT Exp	0.87	1.11	1.16	1.11	2.16	1.72
Industrial	4.03	3.63	3.06	2.74	4.46	3.02
GIDC	2.20	1.93	1.62	1.65	2.94	3.00
Urban	19.40	18.07	15.92	15.43	15.13	15.10
Jyotigram	49.06	46.39	43.25	42.69	41.81	41.95
Ag.	25.72	31.22	16.56	14.72	23.23	23.55
Overall	20.58	21.93	15.73	14.97	17.72	17.67
Overall excluding Ag	18.06	16.77	15.37	15.06	15.40	15.60

Table 2 Overall % POWER Losses OF GUVNL

Category	2008-2009	2009-2010	2010-2011	April-10 to Jan-11	April-11to Jan-12	Modified % as on Jan-12
Distri. Losses	20.58	21.93	15.73	14.97	17.72	17.67
Transm.Losses	4.20	4.64	4.24	4.45	4.54	4.22

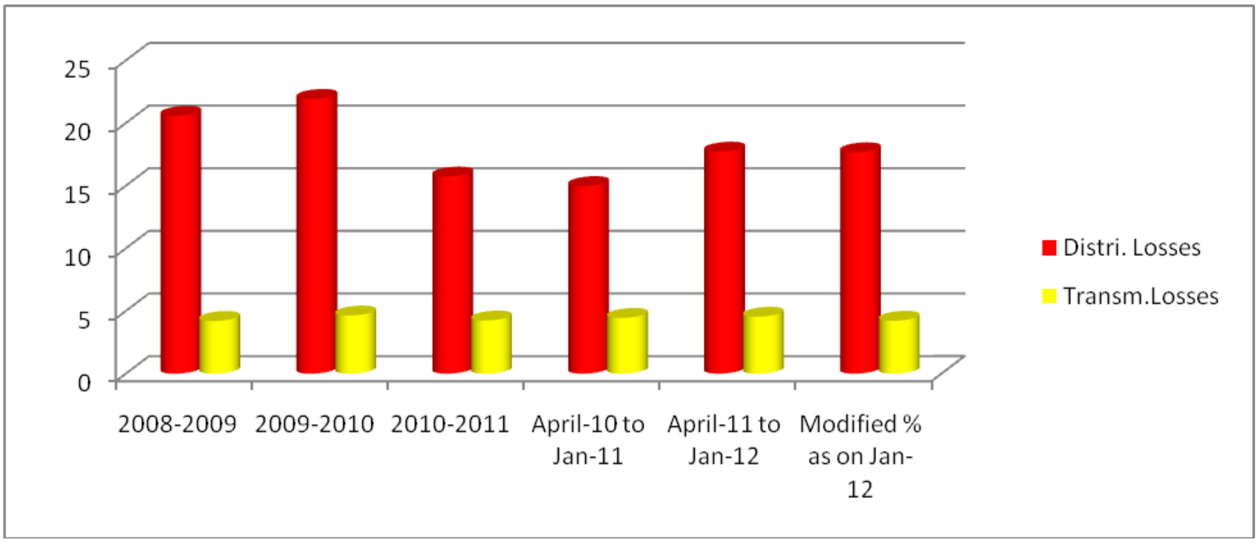

Figure 1 power Losses of GUVNL

II. CAUSES OF DISTRIBUTION POWER LOSES

- **Power theft**

  Defined as a conscience attempt by a person to reduce or eliminate the amount of money he or she will owe the utility for electric energy.
- **Illegal Connection**

  Pure
  Administrative
  Self Line Constructions
- **Meter related problems**

  Meter bypass/ tamper
  Wrong readings
  "U- TOUCH"
  High meter position
  Faulty meter
  Meter ratings
  Over aged meters
  Incorrect C.T. multiplying factor
  Meter Transfers
- Billing Problems
- Premises connected but a/cs not set up in System.
- Wrong classification
- Under keying of meter reading
- Customers with zero readings
- Over clocked meter readings
- Un metered/flat rate supplies

- Under estimation
- Unpaid bills

This paper presents a framework to identify and detect Power Loss activities in the electric utility market i.e., customers with irregular and abnormal consumption patterns indicating fraudulent activities. An automatic feature extraction method for load profiles with a combination of Extreme Learning Machine (ELM), OS-ELM & Support Vector Machines (SVMs) is used to identify fraud customers. This study uses historical customer consumption data collected from GUVNL. Customer consumption patterns are extracted using data mining and statistical techniques, which represent customer load profiles. Based on the assumption that load profiles contain abnormalities when a fraud event occurs, ELM,OS-ELM & SVM classifies load profiles of customers for detection of fraud suspects. There are several different types of fraud that can occur, but our research concentrates only on scenarios where abrupt changes appear in load profiles, indicating fraudulent activities.

I. Proposed Framework Analysis for Power loss

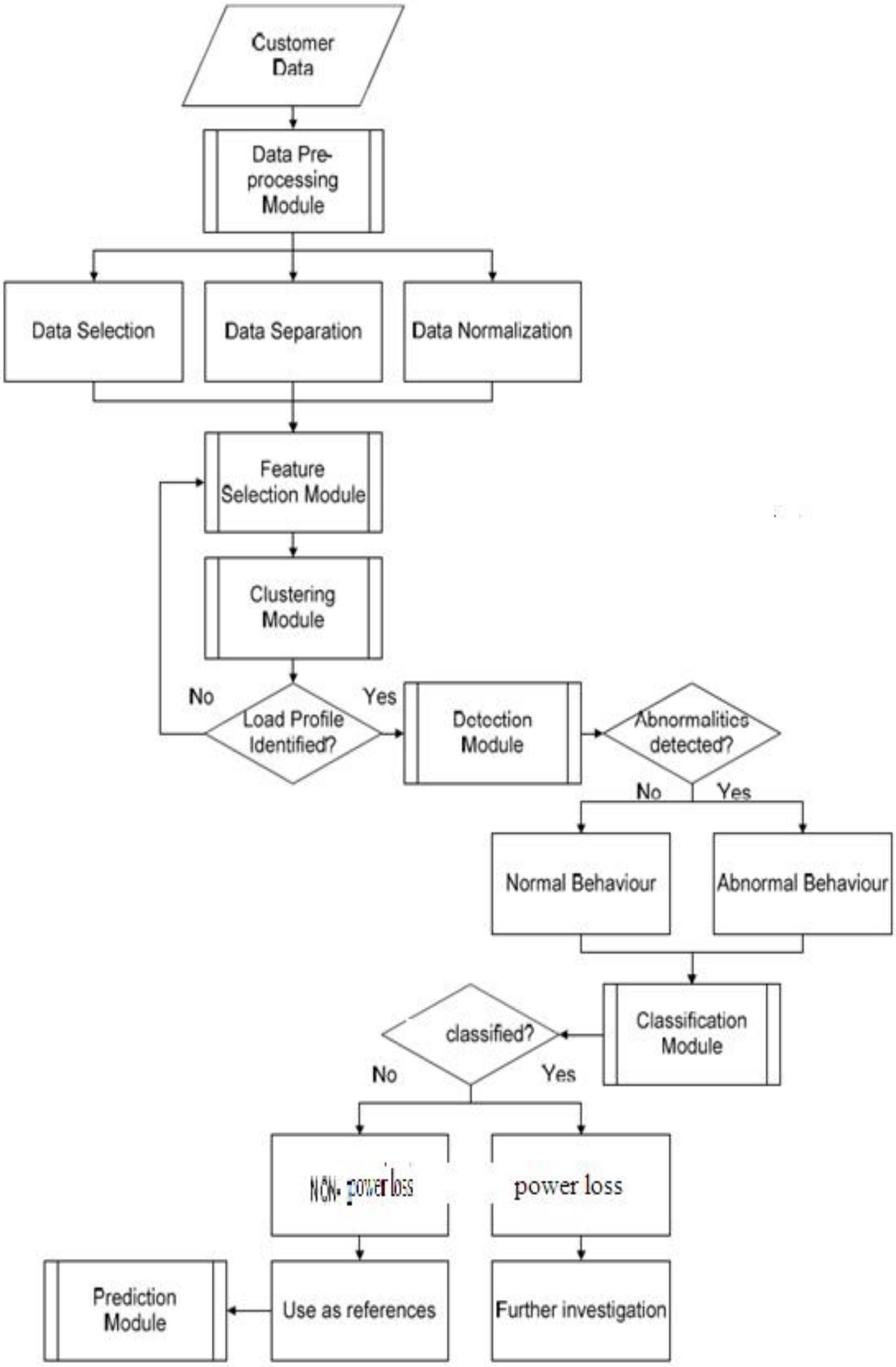

Figure 2 Proposed Framework Analysis for Power loss

Figure 2details the sequence of processes involved in Identifying, detecting, and predicting POWER LOSSs by means of analyzing the customer load consumption database of a power utility, with particular reference to Gujarat GUVNL. The main tasks in this POWER LOSS framework of analysis are summarized below.

Representative load profiles are acquired from the load profiling module as a reference.

Generally, most researchers have rejected untypical load profiles as due to insignificant behavior. However, in the present study, these untypical load profiles are used as benchmarks for investigation by means of outlier detection techniques.

Benchmarks based on two types of load profiles, namely those indicating abnormal behavior patterns and normal behavior patterns are established for use as training data. The abnormal patterns are subsequent Power loss to be investigated further.

Typical load profiles provide reference points in detecting POWER LOSS activities when they are compared with newly emerging load profiles. If no outliers are apparent, then the new load profiles are updated. But if there is some significant deviation, then the new load profiles concerned are investigated using one of the alternative statistical-based, density-based, Distance-based, model-based, or deviation-based outlier detection techniques.

If outlier detection is confirmed as being due to POWER LOSS activity, then the new load profile is updated as an anomalous reference for forecasting purposes. This load profile is then used as a reference for forecasting indications of POWER LOSS activities among new load profiles by applying forecasting techniques such as SVM and time-series analysis. These forecasting techniques are to be explored later below.

The three algorithms ELM, OS-ELM, and SVM have been selected for use in the classification and prediction procedures that are applied to electricity utility customer behavior in the present Context. All these algorithms additionally apply both the sigmoid and RBF nodes activation Functions.

Extreme Learning Machine (ELM)

ELM was proposed by Huang for single, hidden-layer feed-forward neural networks (SLFNs) and it was so devised as to produce superior performance. This algorithm was claimed to be extremely fast in its learning speed and to have better generalization performance when compared to conventional algorithms. Learning speeds that are claimed to be slower due to a slow gradient-based iterative learning algorithm are used extensively to train neural networks. Unlike many other popular learning algorithms, little human involvement is required in ELM. Except for the numbers of the hidden neurons (insensitive to ELM), no other parameters need to be tuned manually by users because this algorithm chooses the input weights randomly and analytically determines the output weights.

Usually, a SLFN has three kinds of input parameters: the input weight *i w*, the hidden neuron biases *i b*, and the output weight *I* b. And, while conventional learning algorithms of SLFNs have to tune these three types of parameters, ELM just randomly generates the input weight *i w* and the hidden neuron biases *i b*, and then analytically calculates the output weight *I* b. No further learning is required for SLFNs trained by ELM. ELM is a general learning algorithm for SLFNs that works effectively for function approximations, classifications, and online prediction problems. Moreover, it can generally work well for a variety of types of applications. Given N arbitrary distinct samples (X_t, t_1), $where\ X_t = [X_{t1}, X_{t2}, \dots, X_{tn}]T \in R^n\ and$ $T_t = [T_{t1}, T_{t2}, \dots, T_{tn}]T \in R^m$, an SLFN with N hidden neurons and activation function g(x) can be mathematically modeled as

$$\sum_{i=1}^{N} B_i g\ (W_i X_j + b_i) = O_j, J = 1_{,\dots\dots,} N$$

Where *i w* is the weight vector connecting the input neurons and the *i*th hidden neuron, *I* b is the weight vector connecting the *i*th hidden neuron and the output neurons, and *I b* is the threshold of the ith hidden neuron. Hear W_i, X_j denotes the inner product of $W_i\ and\ X_j$. If the SLFN can approximate these N sample with Zero error, Then $\sum_{j=1}^{N} ||O_j - t_j|| = 0$ follows; that is, there exist $B_i, W_i, B_i\ such\ that\ \sum_{i=1}^{N} B_i\, g\ (W_i, X_j, +B_i) = t_j, J = 1 \dots\dots N.$

The above N equation can be written compactly as HB=T, where H $(w_1, \dots, w_N, b_1, \dots, b_N, x_1, \dots, x_N) =$

$$\begin{bmatrix} g(w_1 x_1 + b_1) & --- & g(w_N . x_1 + b_N) \\ g(w_1 x_N + b_1) & --- & g(w_N x_N + b_N) \end{bmatrix}$$

N$\times N$

$$\beta = \begin{bmatrix} \beta_1{}^T \\ \beta_N{}^T \end{bmatrix} Nxm \quad \text{and}$$

$$T = \begin{bmatrix} t_1{}^T \\ t_N{}^T \end{bmatrix} Nxm$$

As specified in Huang and Babri, *H* is called the hidden-layer output matrix of the neural network, with the *i*th column of *H* being the *i*th hidden neuron output with respect to inputs *x*1, *x*2,..., *xN*. Based on their previous work, matrix *H* is square and invertible only if the number of hidden neurons is equal to the number of distinct training samples, *N* = *N* ~, indicating, therefore, that SLFNs can approximate these training samples with zero error. In most cases, the number of hidden neurons is much lower than the number of distinct training samples, *N* << *N* ~, *H* is a non square matrix and there may not exist $w_i, b_i \beta_i (i = 1, \dots N)$ such that Hβ = T. Thus, one specific set of $w_i, b_i \beta_i (i = 1, \dots N)$ needs to be found such that

$||$H($w_1, \dots, w_N, b_1, \dots b_N)\ B - T\ || =$

min $||$H$(w_1, \dots, w_N, b_1, \dots b_N) B - T||$

Which is equivalent to minimizing the cost function $E = \sum_{j=1}^{N} (\sum_{j=1}^{N} B_i\ g(w_i . x_j + b_i) - b_i{}^2)$ As Huang maintains, the hidden neuron parameters need not be tuned as the matrix *H* indeed converts the data from nonlinear separable cases to high dimensional linear separable cases. Simulations of numerous real-world datasets (with noise) have shown that ELM performs well in such different cases.

However, Huang showed that the input weights and hidden neurons or kernel parameters are not necessarily tuned and can be randomly selected and then fixed. In fact, the parameters of these hidden neurons are not only independent of each other, but also independent of the training data. Thus, for fixed input weights and the hidden layer biases or kernel parameters, training a SLFN is simply equivalent to finding a least squares solution bˆ for the linear system = *T* . The unique smallest-norm least squares solution for the above linear system is bˆ = *H*†*T*, where *H*† is the Moore-Penrose generalized inverse of the hidden-layer output matrix *H*. As confirmed by Huang, this method may tend to achieve good generalization performance. The ELM algorithm consisting of only three steps can be summarized as set out below. Algorithm ELM given a training set $\aleph = \{(x_1, t_1) | x_1 \epsilon\ R, i = 1, \dots, N\}$, activation function g(x), and hidden neuron number N.

step1: Assign random impur weight w and bias $b_1, i = 1, \dots, N.$
step2: Calculete the hidden- layer output matrix H.
step3: Calulete the output weight B:B=H'T. where H.B. and T are defined as formula (5) and (6).
ELM can accommodate the nonlinear activation function and kernel function. Furthermore, it can avoid difficulties such as stopping criteria, learning rate, learning epochs, and local minima very effectively and in this respect, it is unlike other tuning or adjustment methods.

Huang's analyses established that there are several methods to calculate the Moore-Penrose generalized inverse of a real or complex matrix. These methods may include, but are not limited to, orthogonal projection, orthogonalization method, iterative method, and Singular Value Decomposition (SVD). The orthogonalization method and iterative method have their limitations because searching and iteration are used and these are processes that should be avoided in ELM applications. The orthogonal projection method can be used when *H*H* is non-singular and H† (*H *H*) 1*H** = - . However, *H*H* may not always be non-singular or may tend to be singular only in some applications and, as a consequence, the orthogonal projection method may not perform well in all applications. It is, then, the SVD that can be generally be used to calculate the Moore-Penrose generalized inverse of *H* in all cases. ELM does not require the matrix *H* to be invertible as the SVD method is used in the boosting phase. Moreover, it remains suitable for non-invertible matrices.

ELM provides for a very fast learning capability compared with the alternatives. In, the Performance of the ELM algorithm was compared with other algorithms, including SVM and Conventional back-propagation. The results showed that ELM has an outstanding performance in Relative terms. Furthermore, as Huang and Chen have rece Power loss proven, ELM is actually a Learning algorithm for generalized single hidden-layer feed-forward networks. Besides sigmoid Networks and RBF networks, such SLFNs also include trigonometric networks, threshold networks, fuzzy inference systems, fully complex neural networks, high-order networks, ridge polynomial networks, and wavelet networks.

OS-ELM

In order to handle online applications, the variant of ELM referred to as OS-ELM has been Proposed in. This algorithm was originally developed for single hidden-layer feed-forward Networks with additive or radial basis function (RBF) hidden nodes in a unified framework. The OS-ELM can learn data chunk-by-chunk and that accommodates its application in real-world Industrial contexts. According to, OS-ELM also works for other types of SLFNs that may not be neural networks.

In the derivation of sequential ELM, only the specific matrix *H* is considered, where the rank of *H* is equal to the number of hidden neurons: *rank H N* ~ ( ) =. Under this condition, the following implementation of the pseudo inverse of *H* is readily derived and given by *H* (*HTH*) *HT* † -1 =, which is often called the left pseudo-inverse of *H* from the fact that *N H H I* ~ † = . The corresponding estimation of b is given by B=$(H^T H)^{-1} H^T T$ which is called the least squares solution to *H*b = *T*. The sequential implementation of equation (12) can be derived and is referred to as the recursive least-squares (RLS) algorithm.

In, Liang et al. reviewed the ELM and OS-ELM approaches. OS-ELM consists of two main phases, specifically the boosting phase or initialization phase and the sequential-learning phase. The boosting phase trains the SLFNs using the primitive ELM method given some batch of training data in the initialization stage. These data are discarded once the process is completed. After that, the OS-ELM will learn the training data chunk-by-chunk. Subsequent Power loss, all the training data will be discarded once the learning procedure involving these data is completed. The work done by these authors is set out below in summary form.

Algorithm OS-ELM: Given an activation function g or RBF kernel f , and hidden neuron or RBF kernel number Ñ for a specific application, the following two steps taken.

Step 1: boosting phase: Given a small initial training set ℵ = $\{(x_1 . t_1) | x_1 \epsilon R^m, t_1 \epsilon R^m, i = 1, \dots . N\}$, the intention is to boost the learning algorithm by means of the following procedure:
Assign random input weight w_1 and bias $b_1$or center μ_1 and impact width $\sigma_1, i = 1, \dots, N.$
calculate the initial hidden-layer output matrix $H_0 = [h_1, \dots h_N]T, where\ h_i = [(w_1 x_1 + b_1), \dots, g(w_N x_1 + b_N)]T, i = 1, \dots, N.$

Estimate the initial output weight $B^{(\sigma)} = M_0 M^T{}_0 T_0 where\ M_0 = (H_0{}^T T_0{}^{-1})$
and $T_0 = [t_1, \dots, t_N]T.$
Set k = o.
Step 2: Sequential-earning phase: for each further incoming observation (x_1, t_1), where $x_1 \epsilon R^m, t_1 \epsilon R^m\ and\ i = N + 1. N + 2, N + 3, \dots,$
do the following:
Calculate the hidden-layer output vector $h_{K+1} = [(w_1 x_1 + b_1), \dots, g(w_N x_1 + b_N)]T.$

Calculate latest output weight B^{K-1} based on a Recursive least square (RLS) algorithm:

$$M_{K+1} = M_k - \frac{M_k h(k+1) h^t k+1\ M_k}{1 + h^t\ K+1\ M_k\ h_{k+1}}$$

$$B^{(k+1)} = B^{(k)} + M_{k+1} h_{k+1} (tT^i - hT_{k+1} B^{(k)}$$

Set k=k+1.

ELM and OS-ELM have been selected as the main classifiers for the POWER LOSS analysis to be undertaken in the present research. Both algorithms were claimed to have superior performance including to be extremely fast in its learning speed and to have better generalization performance when compared to conventional algorithms. Unlike many other popular learning algorithms, little human involvement is required in ELM. Except for the numbers of the hidden neurons (insensitive to ELM), no other parameters need to be tuned manually by users because this algorithm chooses the input weights randomly and analytically determines the output weights. Based on the above advantages, so far it is hard to point the limitation exists. In chapters 6 and 7 below, their performance will also be compared with that of the SVM approach to such analysis for the purpose of validation.

Support Vector Machine (SVM)

SVM has emerged as one of the most popular and useful techniques for data classification. The objective of SVM is to produce a model that predicts the target value of data instances in the testing set in which only attributes are given. For the sake of completeness, the fundamentals of the SVM approach are reviewed briefly here.

The goal of an SVM is to estimate a function that is as close as possible to the target values *i y* for

Every *i x* and, at the same time, is as flat as possible for good generalization. Given a set of data

Points $(X_1,Y_1),(X_2,Y_2),\ldots\ldots\ldots(X_1,Y_1)$, where $X_i \in X \in R^n, Y_i \in R$ and L is the total number of training sample, SVM approximates the function using the for lowing form $f(x) = w.\emptyset(x) + b$

Where f (*x*) represents the high dimensional feature spaces which are non-linearly mapped from the input space *x*. The coefficients *w*and *b* is estimated by minimizing the following regularized risk function:

$$\frac{1}{2}||W||^2 + \frac{C}{l}\sum_{i=1}^{l}(\varepsilon_i + \varepsilon_i)$$

There are two types of SVM, Namely the linear SVM and the nonlinear SVM. The linear SVM can occur in two cases, one involving separable data and the other involving non-separable data. The simplest form of SVM classification is the maximal margin classifier. In a linear machine trained on separable data with training data $\{X_i, Y_i\}, i = 1,\ldots\ldots Y_i \in \{-1,1\}, X_i, \in R^d,$ suppose the hyper-plane that separates the positive and negative examples satisfies w.x+b=0. Where W is normal to the hyper-plane, b/|W| is the perpendicular distance from the hyper-plane to the origin, and ||W|| is the Euclidean norm of W. This can be formulated as

$$X_i, W + b \geq +1 \; for \; y_i = +1$$

$$X_i, W + b \geq -1 \; for \; y_i = -1$$

and can the be combined into one set of inequalities.

The maximal margin is $\gamma = \frac{(X_1,W)+b}{|W|} - \frac{(X_2,W)+b}{|W|} = \frac{2}{|W|}$

Therefore, the maximal margin classifier problem can be written in the following form: minimize $\frac{|W|^2}{2}$ $subject\ to\ Y_i\left(x_i, W + b\right) \geq 1, i = 1,\ldots\ldots l$

The Lagrange multiplier method can be used to solve this optimization problem. In most real world problems, the training data are not linearly separable. There are two methods to modify the linear SVM classification to suit the non-linear case. The first is to introduce some slack variables to tolerate some training errors so as to decrease the influence of noise in the training data. This classifier with slack variables is called a soft margin classifier.

Consider the training data

$\{(X_1, Y_1)\ldots\ldots(X_1,Y_1)\}, x \in R^n, Y \in \{-1,+1\}$

with the assumption that they are linearly separable. That is, there exists a hyper-plane

<W, X> + b=0

That satisfies the following constraints: for every $(X_i,Y_i), i = 1\ldots., l \;\; Y_i(< W, X_i >$+b)>0, where <W, X> is the dot product between W and X.

The aim of the maximal margin classifier is to find the hyper-plane with the largest margin, that is, the maximal hyper-plane. This problem can be represented as:

Minimize $\frac{|W|^2}{2}$

Subject to

$Y_i, (< W, X, > +b) \geq 1, i = 1\ldots., l$

In real-world problems, training data are not always linearly separable. In order to handle the non-linearly separable cases, some slack variable have been introduced into the SVM so as to tolerate.

Any training errors, with the influence of the noise in training data thereby decreased. This classifier with slack variables is referred to as a soft-margin classifier. A redial basis kernel function used in this experiment is of the form

$$K\ (X,Y) = exp\ (-\frac{|X - Y|.^2}{2\sigma^2}$$

II. Conclusion

This Paper presents the proposed POWER LOSS analytical framework and the associated key data-mining algorithms to be used in POWER LOSS activity identification, detection, and prediction, including Extreme Learning Machine (ELM), Online Sequential ELM (OS-ELM), and Support Vector Machine (SVM). Several key flow processes in pre-labeling customers' behavior are also described in this, along with a pre-processing procedure to separate the raw data on the basis of types of days so as to normalize the data. This chapter also presents customers' load analysis and its assesses the correlation of such loads with time, weather, and calendar-events factors.

REFERENCES

[1] 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 T&D Conference and Exhibition 2002: Asia Pacific, Vol. 3, pp. 2251-2256.

[2] C. R. Paul, "System loss in a Metropolitan utility network" *IEEE Power Engineering Journal*, pp. 305-307, Sept. 1987.

[3] 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.

[4] Annual Report Tenaga Nasional Berhad 2004, TNB 2004.

[5] T. B. Smith, "Electricity theft: A comparative analysis" *Energy Policy*, Vol. 32, pp. 2067-2076, 2004.

[6] M. S. Alam, E. Kabir, M. M. Rahman, and M. A. K. Chowdhury, "Power sector reform in Bangladesh: Electricity distribution system" *Energy*, Vol. 29, pp. 1773-1783, 2004.

[7] A. Kumar and D. D. Saxena, "Decision priorities and scenarios for minimizing electrical power loss in an India power system network" *Electrical Power Components and Systems*, Vol. 31, pp. 717-727, 2003.

[8] M. A. Ram and M. Shrestha, "Environmental and utility planning implications of electricity loss reduction in a developing country: A comparative study of technical options" *International Journal of Energy Research*, Vol. 22, pp 47-59.

[9] A. H. Nizar, Z. Y. Dong, and Y. Wang, "Power Utility Nontechnical Loss Analysis With Extreme Learning Machine Model" *IEEE Trans. on Power Systems*, Vol. 23, No. 3, pp. 946-955, August 2008.

[10] 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, Sept. 2004.

[11] A. H. Nizar, Z. Y. Dong, J. H. Zhao, P. Zhang, "*A Data Mining Based NTL Analysis Method*" IEEE PES General Meeting, pp. 1-8, 2007.

[12] 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 12th Conference on the Electric Power Supply Industry, Nov 1998.

[13] A. H. Nizar, Z. Y. Dong, M. Jalaluddin, and M. J. Raffles, "*Load Profiling Non-Technical Loss Activities in a Power Utility*" in Proc. of First International Power and Energy Conference (PECON), Nov. 2006.

[14] D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "Allocation of the load profiles to consumers using probabilistic neural networks" *IEEE Trans.on Power Systems*, Vol. 20, No. 2, pp. 548-555, May 2005.

[15] V. Vapnik. *Statistical Learning Theory*, John Wiley & Sons, 1998.

[16] D. Wang, D. S. Yeung, E. C. C. Tsang, "Weighted Mahalanobis Distance Kernels for Support Vector Machines" *IEEE Trans. on Neural Networks*, Vol. 18, No. 5, pp. 1453-1462, Sept. 2007.

[17] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

[18] C.-W. Hsu, C.-C. Chang, and C.-J. Lin. A practical guide to support vector classification. Technical report, Dept. of Computer Science, National Taiwan University, 2003.

[19] 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.

[20] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: Michigan Univ. Press, 1975 (Cambridge, MA: MIT Press, 1992).

[21] 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.

[22] A. H. Nizar, and Z. Y. Dong, "Identification and detection of electricity customer behavior irregularities," in *Proc. of IEEE/PES Power Systems Conference and Exposition*, Seattle, Washington, United States, Mar. 15-18, 2009, pp. 1–10.

[23] A.H. Nizar, J.H. Zhao, and Z.Y. Dong, "Customer Information System Data Pre-Processing with Feature Selection Techniques for Non-Technical Losses Prediction in an Electricity Market," in Proc. of the International Conference on Power System Technology (PowerCon) 2006, Chongqing, China, Oct. 22-26 2006, pp. 1–7.

[24] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines", *IEEE Trans. Power Del.*, vol. 25, no. 2, pp. 1162–1171, Apr. 2010.

[25] C. R. Paul, "System loss in a Metropolitan utility network" *IEEE Power Engineering Journal*, pp. 305–307, Sept. 1987.

[26] J. Nagi, K.S. Yap, S.K. Tiong, and S.K. Ahmed, "Detection of Abnormalities and Electricity Theft using Genetic Support Vector Machines", in Proc. of the IEEE Region 10 Conference (TENCON) 2008, Nov. 18-21, 2008, Hyderabad, India, pp. 1–6.

[27] A. H. Nizar, Z. Y. Dong, J. H. Zhao, P. Zhang, "A Data Mining Based NTL Analysis Method", IEEE PES General Meeting, pp. 1–8, 2007.

[28] 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 T&D Conference and Exhibition 2002: Asia Pacific, Vol. 3, pp. 2251–2256.
D. Gerbec, S. Gasperic, I. Smon, and F. Gubina, "Allocation of the load profiles to consumers using probabilistic neural networks" *IEEE Trans. on Power Systems*, vol. 20, no. 2, pp. 548–555, May 2005.

[29] Hydro and Thermal Power Business Team, Transmission and Distribution Loss Reduction Services #167 Samsung-dong, Kangnamgu, Seoul, 135-791, Korea.

[30] H.L. Pok, K.S. Yap, I.Z. Abidin, A.H. Hashim, Z.F. Hussien, and A.M. Mohamad, "Abnormalities and Fraud Electric Meter Detection using Hybrid Support Vector Machine and Modified Genetic Algorithm", in Proc. of the 19th International Conference on Electricity Distribution (CIRED) 2007, Vienna, May 21–24 2007.

[31] M. Kantardzic, "Data Mining: Concepts, Models, Methods and Algorithms", Ed. IEEE Press,2003.

[32] G. Piatetski-Shapiro, W.J. Frawley, "Knowledge discovery in databases", Ed. AAAI/MIT Press, 1991.

[33] Yufeng Kou, Chang-Tien Lu, Sirirat Sinvongwattana, Yo-Ping Huang, "Survey of Fraud Detection Techniques", Proceedings of the 2004 IEEE International Conference on Networking, Sensing & Control, Taiwan, March 21, 2004.

[34] J.R. Galván, A. Elices, A. Muñoz, T. Czernichow, M.A. Sanz-Bobi, "System for Detection of Abnormalities and Fraud in Customer Consumption", 12th Conference on the Electric Power Supply Industry, November,1998, Thailand.
[35] R. Wheeler, S. Aitken, "Multiple Algorithms for Fraud Detection", Knowledge-Based Systems 13 (2000) 93–99.