Support Vector Machine Approach for Non-Technical Losses Identification in Power Distribution Systems

Kalyan Pal
Dept of Electrical Engineering
ABES Institute of Technology, APJAKTU
Ghaziabad, India
e-mail: kalyanpal@yahoo.com

Bhavesh Kumar Chauhan
Dept of Electrical Engineering
ABES Institute of Technology, APJAKTU
Ghaziabad, India

Abstract: Electricity consumer fraud 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. In this paper,the approach towards nontechnical loss (NTL) detection in power utilities using an artificial intelligence based technique, Support Vector Machine (SVM), are presented. This approach provides a method of data mining, which involves feature extraction from past consumption data. This SVM based approach uses customer load profile information and additional attributes to expose abnormal behavior that is known to be highly correlated with NTL activities. Some key advantages of SVM in data clustering, among which is the easy way of using them to fit the data of a wide range of features are discussed here. Finally, some major weakness of using SVM in clustering for NTL identification are identified, which leads to motivate for the scope of Optimum-Path Forest, a new model of NTL identification.

Keywords: *Non-Technical Losses, Support Vector Machine, Clustering, Optimum-Path Forest*

*****

I. INTRODUCTION

The Non-Technical Losses (NTL) refers to the energy consumed illegally without paying the bill. These lead to an uncountable situation between the registered consumption to the distribution company and the real consumption of the consumers. This creates important economic losses for the utility company. Usually, the detection of NTLs use features such as the customers' consumption, economic activity, geographic location, the contracted power and the active/reactive ratio. However, distribution companies have lots of information: documentation, inspectors' commentaries, additional information about customers' facility, etc in their database. At present, there are no references that use specific techniques to treat the additional information which can be identified in the company database. Usually, they are confined to use consumption data and limited information related to the customer.

Data processing in large datasets has been a challenge for the machine-learning community. There are various methods, that have been proposed to overcome the problem of overlapping samples in a very effective manner, their computational burden for training may be prohibitive in applications that require retraining at every time step.Alternatively, considerable efforthasbeen dedicated to developing feature selection techniques in order to speed up the classification process and to increase its accuracy. Clustering is one of methods for analyzing and processing large and not well-known amount of data [1]. This is the method of classifying the data set into subsets, clusters, based on a defined similarity measure. On this way, a set of characteristic states that describe analyzed problem can be generated.

Patterns are usually represented by feature vectors (set of measures or observations) obtained from samples of a dataset. Two fundamental problems in pattern recognition are: (i) the identification of natural groups (clustering) composed by samples with similar patterns and (ii) the classification of each sample in one of the possible classes (labels). The dataset is usually divided in two parts, a training set and a test set, being the first used to project the classifier and the second used for validation, by measuring its classification errors (accuracy). While problem (i) has no prior information about the labels of the samples, the training in problem (ii) can count with unlabeled samples (unsupervised learning), labeled samples (supervised learning) or part of the samples labeled and the other part unlabeled (semi-supervised learning) [2].

There are three typical cases in 2D feature spaces using two classes: (a) linearly separable, (b) piecewise linearly separable, and (c) non-separable classes with arbitrary shapes. Any reasonable approach should handle (a) and (b), being (c) the most interesting challenge. An artificial neural network with multi-layer perceptrons (ANN- MLP), for example, can address (a) and (b), but not (c) [3]. As an unstable classifier, collections of ANN-MLP can improve its performance up to some unknown limit of classifiers. Support vector machines (SVMs) have been proposed to overcome the problem, by assuming linearly separable classes in a higher-dimensional feature space. Its computational cost rapidly increases with the training set size and the number of support vectors. As a binary classifier, multiple SVMs are required to solve a multiclass problem. Tang and Mazzoni [4]proposed a method to reduce the number of support vectors in the multi-class problem.

ISSN: 2321-8169
158 – 162

Their approach suffers from slow convergence and high computational cost, because they first minimize the number of support vectors in several binary SVMs, and then share these vectors among the machines. Another point is that, in all SVM approaches, the assumption of separability may also not be valid in any space of finite dimension.

Support Vector Machine (SVM), another approach towards nontechnical loss (NTL) detection in power utilities using an artificial intelligence based technique, were introduced by Vapnik in the late 1960s. The SVM, based on the foundation of statistical learning theory, is a general classification method. SVMs have recently been applied to several applications ranging from face identification, text categorization to Bioinformatics, and database mining [5].

The Support Vector Machines (SVMs) are usually considered the first practical application of statistical learning theory. This is a research area that offers many options to work, most of them being more conceptual than merely technical. In recent years, its scope has increased significantly both in terms of new algorithms and a further theoretical understanding. Part of these new algorithms is due to kernelizing methods, a proposal for solution of problems of machine learning whose architecture has proved able to handle issues relating to the basis of this theory. Moreover, successful applications of SVMs have shown that this technique not only has a more solid substantiation as Artificial Neural Networks but are also able to replace them with similar or better performance [6].

The fraud detection using SVM method is very promising, as this method achieves the highest inspection hit-rate for fraud customer detection. Firstly, it should be noted that, SVM has non-linear dividing hyper surfaces that give it high discrimination. Secondly, SVM provides good generalization ability for unseen data classification. Lastly, SVM determines the optimal network structure itself, without requiring to fine tune any external parameters, as in the case of the BPNN and the OS-ELM. In contrast to the advantages of SVM over neural networks, there are however some drawbacks of SVM. These drawbacks are restricted due to practical aspects concerning memory limitation and real-time training. Some of the major drawbacks of SVM are as follows. The optimization problem arising in this method is not easy to solve. Since the number of Lagrange multipliers is equal to the number of training samples, the training process is relatively slow. Even with the use of the SMO algorithm, real time training is not possible for a large set of data [7].

II. THE SVM MODEL

A model is produced by the SVM to predict the target value of data instances in the testing set in which only attributes are given. The goal in SVM in a binary classification problem is to separate two classes by means of a function devised from available data and hence to produce a classifier that should work well with further unseen data. For the sake of completeness, the fundamentals of SVM approach are briefly reviewed here.

SVM classification is the simplest form to maximize the margin classifier. It is used to solve the most basic classification problem, namely the case of a binary classification with linear separable training data. For a linear separable dataset, a linear classification function corresponds to a separating hyperplane f (x) that passes through the middle of the two classes, separating the two. Once this function is determined, new data instance x_n can be classified by simply testing the sign of the function f (x_n); x_n belongs to the positive class if f (x_n)> 0.

The main purpose of the (binary) SVM algorithm used for classification is to construct an optimal decision function, which accurately predicts unseen data into two classes and minimizes the classification error using

$$f(x) = sgn(g(x)) \quad (1)$$

here $g(x)$ is the decision boundary between the two classes.

This is achieved by following the method of structural risk minimization (SRM) principle, given by [7]

$$R < \frac{t}{n} + \sqrt{\frac{h\left(\ln\left(\frac{2N}{h}\right)+1\right) - \ln\left(\frac{\eta}{4}\right)}{n}} \quad (2)$$

Where ‘R’ is the classification error expectations’ is the number of training errors, ‘n’ is the number of training samples and ‘η’ is a confidence measure.

In the case of separable data, the first term in (2) is zero and the second term is minimized resulting in a good generalization performance of the SVM. The function g(x) in (1) is the decision boundary, which is derived from a set of training samples

$$X = \{x_1, x_2, \ldots, x_n\}, x_i \in \mathrm{R}^M, i = 1, 2, \ldots, n \quad (3)$$

Where each training sample x_i has M features describing a signature and belongs to one of two classes

$$Y = \{y_1, y_{2,\ldots,} y_n\}, y_i \in \{-1, +1\}, i = 1, 2, \ldots, n \quad (4)$$

The decision boundary between the two classes is a hyper-plane described by the equation

$$g(x) = \langle w, x \rangle + b \quad (5)$$

where $'w'$ and $'b'$ are derived in such a way that the unseen data is classified correctly. This isachieved by maximizing the margin of separation between the two classes.

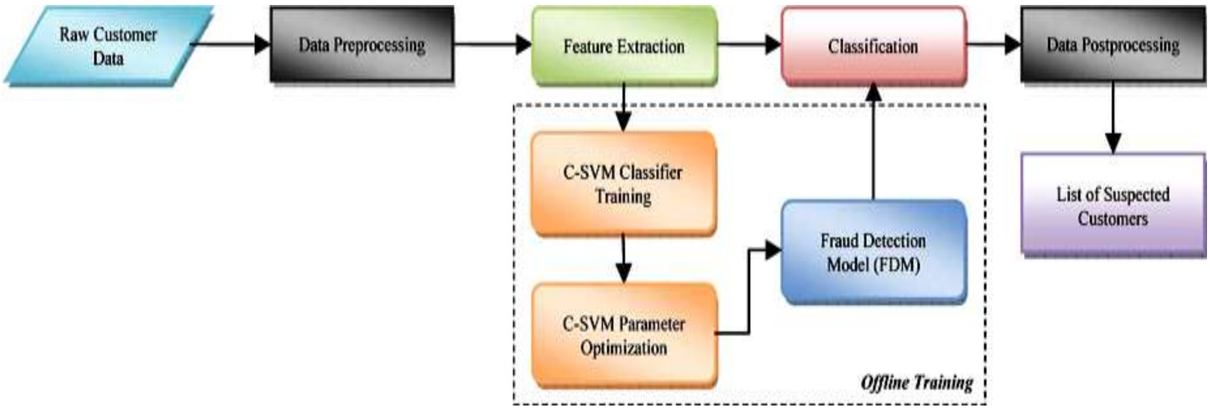

Figure 1. Fraud detection framework for the detection of customers with abnormalities and fraud activities.

An automatic feature extraction method using load profiles with the combination of support vector machine (SVM) is used to identify customers with abnormalities and fraud activities. Customer consumption patterns are extracted using data mining and statistical techniques, which represent load profiles. Based on the assumption that load profiles contain irregularities when a fraud event occurs, SVM classifies load profiles of customers into two categories: normal and fraud [9]. The SVM, based on the foundation of statistical learning theory, is a general classification method.

A Support Vector Machine (SVM) is a maximum margin classifier, i.e. it creates a maximum separation between classes. Therefore, a SVM is less prone to overfitting than other classifiers, such as a neural network. Support vectors hold up the separating hyperplane. In practice, they are just a small fraction of the training examples. The training of a SVM can be defined as a Lagrangian dual problem having a convex cost function. By default, the separating hyperplane is linear. For complex problems, it is advantageous to map the data set to a higher dimension space, where it is possible to separate them using a linear hyperplane [10].

Because there are many such linear hyperplanes, what SVM additionally guarantee is that the best such function is found by maximizing the margin between the two classes. Intuitively, the margin is defined as the amount of space, or separation between the two classes as defined by the hyperplane. Geometrically, the margin corresponds to the shortest distance between the closest data points to a point on the hyperplane. Having this geometric definition allows us to explore how to maximize the margin, so that even though there are an infinite number of hyperplanes, only a few qualify as the solution to SVM. The reason why SVM insists on finding the maximum margin hyperplanes is that it offers the best generalization ability. It allows not only the best classification performance (e.g., accuracy) on the training data, but also leaves much room for the correct classification of the future data [11].

III. LIMITATIONS OF SVM MODEL

Requirement of sufficient storage capacity is the main drawback of SVM algorithm. The support vectors (SVs) represent the important training samples describing the distinguishing features of the given classes. When the optimization problem has a low separability in the space used, the number of SVs increases. These SVs have to be stored in a model file for future classification. This puts limitations on the use of SVM for pattern recognition or classification in devices with limited storage capacity [12].

However, many of these traditional pattern recognition techniques may not be suitable to handle huge volumes of data in real time. Artificial Neural Networks and Support Vector Machines require high computational burden on training. The parameter optimization of the latter technique turns the quadratic optimization problem as an exponential one, which can be a serious problem in case of real time training systems [13].

Popular approaches, such as Support Vector Machines(SVMs) and Artificial Neural Networks (ANNs), present a prohibitive computational time for large datasets, especially in the training phase. Although there have been efforts, such as LASVM and SVMs without kernel mapping,to speed up SVMs, for instance, LASVM is limited to binary classification and SVMs considerably reduces the accuracy of classification in the case of overlapped classes. Therefore, it seems paramount to develop more efficient and effective pattern recognition methods for large datasets [14].

Well known methods, such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN), present high computational cost for training, being impractical in the case of training sets with thousands of samples or in applications that require multiple retraining with interactive response times (e.g., interactive segmentation. In the case of redundant training sets, it is still possible to reduce them to improve efficiency of these approaches. However, reduced training sets usually affect the efficacy of them in a significant way [14].

As an unstable classifier, collections of ANN-MLP can improve its performance up to some unknown limit of classifiers. Support vector machines have been proposed to overcome the problem, by assuming linearly separable classes in a higher-dimensional feature space. Its computational cost rapidly increases with the training set size and the number of support vectors. As a binary classifier, multiple SVMs are required to solve a multi-class problem [15].

An SVM is a type of large-margin classifier: it is a vector space based machine learning technique where the objective is to find a decision frame between two classes that is maximally far from any point in the training data.Support vector machines are not necessarily better than other machine learning techniques (except perhaps in situations with little training data), but they perform at the state-of-art level and have much theoretical and practical appeal.

IV. OPTIMUM-PATH FOREST

Recently, a different framework was introduced for graph-based classifiers that reduces the pattern recognition problem to an optimum-path forest computation (OPF) in the feature space induced by that graph [16]. This kind of classifier does not interpret the classification task as a hyperplane optimization problem, but rather as a combinatorial optimum-path computation based on certain key samples (prototypes) to the remaining nodes.

Such classifiers will not obstruct the tasks related to classification, which is known as the hyper planes optimization issues whereas, the combinatorial optimum-path computation is present for the remaining nodes. Each node is categories as per the strength of the prototypes that are connected which is defined as the discrete optimal partition for the feature space.

Each pattern becomes a root for its optimum-path tree, and each node is classified according to the strength of its connection to the pattern, which defines a discrete optimal partition (influence region) of the feature space. There is a technique proposed by Ramos, C. C. O et al called as a technique which is based on the OPF to identify and to safeguard the NTLs [17]. It is determined that the OPF has higher chances to identify the electricity theft. On the other hand, they are even higher in accuracy when compared to SVM [18].

OPF-based classifiers have certain advantages: i) they are free of parameters, ii) they do not assume any special shape or separability of the feature space, and iii) they run the training phase faster so real-time applications for fraud detection in electrical systems are feasible [19].

V. CONCLUSION

At present in machine learning applications, support vector machines (SVM) are considered one of the most robust and accurate methods among all popular algorithms. It has a stable theoretical foundation, requires only a dozen examples for training, and is insensitive to the number of dimensions. Over and above, efficient methods for training SVM are also being developed at a fast pace. The aim of SVM is to find the best classification function to distinguish between members of the two classes in the training data. The metric for the approach of the "best" classification function can be realized geometrically.

REFERENCE

[1] A.C. Edison, A. Neto, Jorge Coelho, "Probabilistic methodology for Technical and Non-Technical Losses estimation in distribution system", Elsevier, Electric Power Systems Research, vol.97, pp. 93-99, April 2013.

[2] J.P.Papa, A.X. Falcao, V.H.C. de Albuquerque, J.M.R.S. Tavares, "Efficient Supervised Optimum-Path Forest Classification for Large Datasets", Elsevier, Pattern Recognition, vol. 45, pp. 512–520, 2012.

[3] S. Haykin, "Neural Networks:A Comprehensive Foundation", Prentice Hall, 1994.

[4] B. Tang, D. Mazzoni, "Multiclass Reduced-Set Support Vector Machines", In Proceedings of the 23th International Conference on Machine learning, ACM Press, New York, USA, pp. 921–928, 2006.

[5] V.N. Vapnik, "Statistical Learning Theory", New York, Wiley, 1998.

[6] N. Cristianini, J.S. Taylor, "An Introduction to Support Vector Machines", Cambridge, MA: Cambridge Univ. Press, 2000.

[7] G. Dror, R. Sorek, S. Shamir, "Accurate Identification of Alternatively Spliced Exons Using Support Vector Machine", Bioinformatics, vol. 21, no. 7, pp. 897–901, April 2005.

[8] M.E. Mavroforakis, S. Theodoridis, "A Geometric Approach to Support Vector Machine (SVM) Classification", IEEE Trans. Neural Network, vol. 17, no. 3, pp. 671–682, May 2006.

[9] V. Cherkassky, Y. Ma, "Practical Selection of SVM Parameters and Noise Estimation for SVM Regression", Neural Network, vol. 17, pp. 113–126, 2004.

[10] S. Amari, S.Wu, "Improving Support Vector Machine Classifiers by Modifying Kernel Functions", Neural Network, vol. 12, pp. 783–789, 1999.

[11] M.A. Oskoei, H. Hu, "Support Vector Machine-Based Classification Scheme for Myoelectric Control Applied to Upper Limb", IEEE Trans. Biomed. Eng., vol. 55, no. 8, pp. 1956–1965, August 2008.

[12] C. Ramos, A. Souza, J. Papa, A. X. Falcao, "Fast Non-Technical Losses Identification Through Optimum-Path Forest", in Proceedings of 15th International Coreference on Intelligent System Applications to Power Systems (lSAP), pp. 1-5, 2009.

ISSN: 2321-8169
158 – 162

[13] A. Falcao, J. Stolfi, R. Lotufo, “The Image Foresting Transform Theory, Algorithms, and Applications”, IEEE Transactions on Pattern Analysisand Machine Intelligence, vol. 26, no. 1, pp. 19–29, Jan. 2004.

[14] J.P. Papa, A.X. Falcao, V.H.C. de Albuquerque, J.M.R.S. Tavares, “Efficient Supervised Optimum-Path Forest Classification for Large Datasets”, Elsevier, Pattern Recognition, vol. 45, pp. 512–520, 2012.

[15] G.H. Rosa, K.A.P. Costa, Leandro A.P. Junior, J.P. Papa, A.X. Falcao, “On the Training of Artificial Neural Networks with Radial Basis Function Using Optimum-Path Forest Clustering”, 22ndIEEE International Conference on Pattern Recognition, pp. 1472-1477, 2014.

[16] J.P.Papa, A.X. Falcao, C. Suzuki, “Supervised Pattern Classification Based on Optimum-Path Forest”, Int. J. Image. Syst. Technol., vol. 19, no. 2, pp. 120–131, 2009.

[17] C.C.O. Ramos, A.N. de Sousa, J.P. Papa, A.X. Falcao, “Learning to Identify Non-Technical Losses with Optimum-Path Forest”, In: Proceedings of the IWSSIP-2010-17th international conference on systems, signals and image processing, pp. 1–4, 2010.

[18] I. Monederon, F. Biscarri, C. Leon, J. Biscarri, R. Millán., “Midas: Detection of Non-Technical Losses in Electrical Consumption Using Neural Networks and Statistical Techniques”, In: Proceedings of the international conference on computational science and applications, vol. 3984, pp. 725-734, 2006.

[19] C.C.O. Ramos, A.N. de Sousa, J.P. Papa, A.X. Falcao, “A New Approach for Non-Technical Losses Detection Based on Optimum-Path Forest”, IEEE Transactions on Power Systems, Vol. 26, No. 1, February 2011.