International Journal of Soft Computing 2 (4): 544-548, 2007

Categorizing Power System Stability Using Clustering Based Support Vector Machines

[1]B. Dora Arul Selvi and [2]N. Kamaraj
[1]Sivanthi Aditanar College of Engineering, Tiruchendur-628 215, Tamilnadu, India
[2]Thiagarajar College of Engineering, Madurai-15, Tamilnadu, India

Abstract: The current deregulation trend and the participation of many players are contributing to the decrease in security margin. This seeks the development of reliable and faster security monitoring methods. Support Vector Machines, a Neural Network Technology has been as presented an important contributor for reaching the goals of online Transient stability assessment. The training complexity of SVM is highly dependent on the size of data set. Since the power systems are of high dimensionality, feature extraction techniques must be implemented to make the application feasible. This study presents a new Clustering Based SVM to identify the stability status of power system. Here we have applied an exclusive clustering algorithm and an overlapping clustering algorithm, which scan the entire data set only once to provide SVM with high quality samples that carry the statistical summaries of the data such that the summaries maximize the benefit of learning the SVM. Transient stability of New England 39 Bus system is assessed by SVM trained with complete input feature set. The aspects of training time and classification accuracy are compared to the results obtained from CB-SVM.This shows that CB-SVM is highly useful for very large data sets while also generating high degree of classification accuracy.

Key words: Support Vector Machines (SVM), neural networks, data clustering, transient stability assessment

INTRODUCTION

With the power system expansion and increase in their complexity, the dynamic security analysis has become a very crucial and complex process. Therefore, the developments of fast and accurate techniques for security evaluation are the need of the hour. In the last few decades, Transient Stability Assessment (TSA) methods of practical use have been developed and they are mainly based on time domain simulation (Padiyar, 2000). The neural network approach has been proposed as a possible alternative for the analytical TSA because of their faster response (Sobajic and Pao, 1989). The nonlinear input /output mapping of Neural Networks can be used to produce a security index that classifies the current operating point as stable or unstable.

Although successfully applied to TSA, the implementation of Back Propagation and Multilayer Perceptrons require extensive training process (Zhou *et al*., 1994; Pao and Sobajic, 1992). Support Vector Machines, a recently introduced non-linear learning based classifier have been a promising method for classification and regression analysis because of their solid mathematical foundation. However despite their prominent properties, they are also not favoured for very large data set. Employing suitable feature extraction technique has alleviated this problem of high input dimensionality (Moulin *et al*., 2004). At the same time feature extraction should offer sensitivity information to identify the input features best suited for maintaining high degree of classification accuracy.

In a real power system with very large data records even multiple scans of the entire data set are too expensive to perform. For assessing the transient stability of power system, this study presents a new method clustering based SVM, designed especially for very large data set. CB-SVM applies two clustering algorithms (Jain *et al*., 1999) to get a finer description of data closer to the decision boundary and coarser description of data away from the decision boundary. Thus CB-SVM tries to generate the best SVM boundary for very large data sets while maintaining high classification accuracy.

SVM OVERVIEW

In machine learning theory, the optimal class boundary function or hypothesis h(x) given a limited number of training data set{(x,y)}(y is the label of x) is considered the one that gives best generalization performance which denotes the performance on “unseen”

Corresponding Author: B. Dora Arul Selvi, Sivanthi Aditanar College of Engineering, Tiruchendur-628 215, Tamilnadu, India

examples rather than training data. The performance on the training data is not regarded as a good evaluation measure for a hypothesis because a hypothesis ends up over fitting when it tries to fit the training data too hard. When a problem is easy to classify and the boundary function is complicated more than it needs to be, the boundary is likely to be over fit. When a problem is hard and the classifier is not powerful enough, the boundary becomes under fit. SVMs are the excellent examples of supervised learning that tries to maximize the generalization by maximizing the margin and also this supports nonlinear separation using advanced kernels, by which SVMs try to avoid over fitting and under fitting (Burges, 1998). The margin in SVMs denotes the distance from the boundary to the closest data in the feature space.

In SVMs, the problem of computing a margin maximized boundary function is specified by the following quadratic programming problem minimize:

$$W(\alpha) = -\sum_{i=1}^{l}\alpha_i + \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l} y_i y_j \alpha_i \alpha_j k(x_i, x_j) \quad (1)$$

Subject to:

$$\sum_{i=1}^{l} y_i \alpha_i = 0 \quad (2)$$

$$\forall_i : 0 \le \alpha_i \le C \quad (3)$$

The number of training data is denoted by l, . is a vector of l variables, where each component . corresponds to a training data (x ,y). C is a soft margin parameter controlling the influence of noise in training data. The kernel k(x,x) for linear boundary function is x. x, a scalar product of two data points. The nonlinear transformation of feature space is performed by replacing k(x,x) with an advanced kernel such as Polynomial kernel

$$\left(x^T x_i + 1\right)^P \quad (4)$$

(or) RBF kernel $$\exp\left(-\frac{1}{2\sigma^2}\|x - x_i\|^2\right) \quad (5)$$

An advanced kernel is a function that operates on the input data but has the effect of computing the scalar product of their images usually in a much higher-dimensional feature space, which allows one to work implicitly with hyper planes in such complex spaces. Figrue 1 shows the architecture of SVM, where the number of units K(x, xi) is determined by the number of Support Vectors (SVs).KF is the kernel function, which can be linear or nonlinear. Here RBF kernel function represented by Eq. 5 is used.

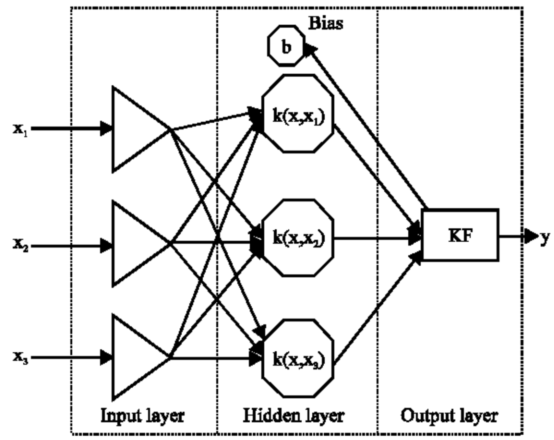

Fig. 1: Architecture of SVM

Another characteristic of SVM is that its boundary function is described by the SVs, which are the data closer to the boundary. The above QP problem computes a vector ., each element of which specifies the weight of each data and the SVs are the data whose corresponding . is greater than zero. In other words, the other data rather than the SVs do not contribute to boundary function and thus computing a SVM boundary function can be viewed as finding the SVs with the corresponding weights to describe the class boundary. There have been many attempts to revise the original QP formulation such that a QP solver can solve it more efficiently (Fung and Mangasarian, 2001). Here we tried to provide a smaller but high quality data set that is beneficial in computing the SVM boundary function effectively by applying two clustering algorithms. Our CB-SVM algorithm substantially reduces the total number of data points for training SVM while trying to keep the high quality of SVs that describes the boundary the best.

DATA CLUSTERING ALGORITHM FOR LARGE DATA SET

Clustering is a process of partitioning or grouping a given set of unlabelled data into a number of clusters such that similar patterns are assigned to one cluster (Veronica, 2002). Each pattern can be represented by a vector having many parameters or attributes. Fundamental to the use of any clustering technique is the measure of similarity or the distance between the respective patterns. Here two most representative algorithms are used.

Exclusive clustering: Here data are grouped in an exclusive way, so that if a certain datum belongs to a definite cluster then it could not be included in another cluster. K-Means clustering algorithm, which is an exclusive data clustering technique, partitions a collection of n vectors x, j = 1…n into C groups, i = 1…C and find a cluster center in each group such that an objective function of dissimilarity is minimized. The Process of K-Means algorithm is as follows.

Step 1: Initialize the cluster center C, i = 1…C. This is typically achieved by randomly selecting C points among all of the data points.

Step 2: Determine the membership function U using the equation

$$U_{ij} = \begin{cases} 1 \text{ if } \left\|x_j - c_i\right\|^2 \le \left\|x_j - c_k\right\|^2 \text{ for, each } k \ne i \\ 0 \text{ otherwise} \end{cases} \tag{6}$$

Step 3: Compute the objective function

$$J = \sum_{i}^{c} J_i = \sum_{i=1}^{c} \left(\sum_{k, x_k \in G_i} \left\|x_k - c_i\right\|^2 \right) \tag{7}$$

Stop if either it is below a certain tolerance value or its improvement over previous iteration is below a certain threshold.

Step 4: Update the cluster center by equation

$$c_i = \frac{1}{|G_i|} \sum_{k, xk \subset Gj} x_k \tag{8}$$

Where
$$|G_i| = \sum_{j=1}^{n} u_{ij} \tag{9}$$

Go to step 2.

Overlapping clustering: Fuzzy C means clustering is an overlapping data clustering algorithm in which each data point belongs to a cluster to a degree specified by membership grade. It employs fuzzy partitioning such that a given data point can belong to several groups with a degree of belongingness between 0 and 1. The process of Fuzzy C Means clustering algorithm is as follows.

Step 1: Initialize the membership matrix u with random values between 0 and 1 such that constraints in equation

$$\sum_{i=1}^{c} u_{ij} = 1, \forall_j = 1, \ldots, n \tag{10}$$

are satisfied.

Step 2: Calculate C fuzzy cluster centers C ,i = 1…C using equation

$$C_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m} \tag{11}$$

Step 3: Compute the objective function

$$J(U, C_1, \ldots, C_c) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \sum_{j}^{n} u_{ij}^m d_{ij}^2 \tag{12}$$

Stop if either it is below a certain tolerance value or its improvement over previous iteration is below a certain threshold.

Step 4: Compute new u by equation.

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(m-1)}} \quad \text{each } k \ne i \tag{13}$$

Go to step2.

RESULTS AND DISCUSSION

To illustrate the applicability of data clustering Techniques connected with SVM (CB-SVM), the New England 39 bus system is used. The Configuration of the system is shown in Fig. 2.The system consists of 10 Generators, 46 Transmission lines and 10 Transformers. The data required for transient stability analysis are given in Appendix A and the data for load flow analysis are obtained from reference (Padiyar, 2000).

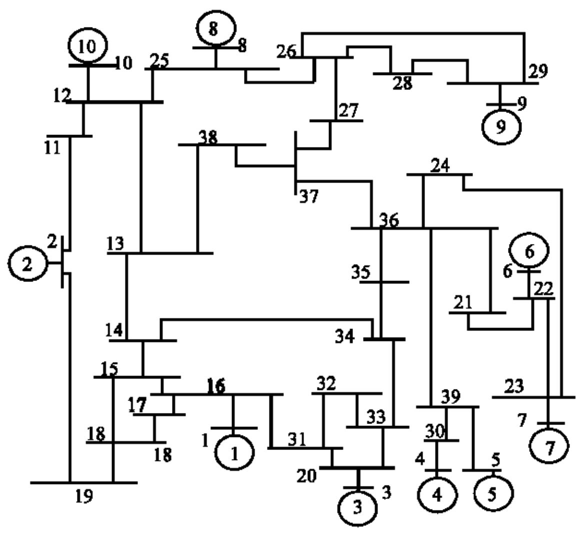

Fig. 2: Configuration of new England 39 bus system

Different operating conditions have been created by changing the load pattern of the system randomly. It is assumed that the contingency set consist of only one disturbance which is a 3 Phase fault on the transmission line connecting bus 14 and bus 34 near bus 14. Duration of disturbance is 0.28 sec. Opening the circuit breakers on both ends of the transmission lines clears the fault. A data set of 600 operating states has been generated and each case has been validated by an optimal power flow execution. Then the contingency has been simulated in time domain. The classification of the system as stable/unstable is determined based on the relative rotor angles of generators in Center Of Inertia (COI) frame of reference. For a given sample if the relative rotor angle of any one machine exceeds 100 degrees then the input state is considered as unstable otherwise it is stable. The complete input feature set consists of the active and reactive powers of each generator, total active and reactive loads of the system at the instant of fault with a total of 22 inputs and one output indicating the security class. Out of the 600 input states generated, 450 datasets are allotted for training and 150 for testing purpose.

SVMs have been trained on the examples with RBF kernel. The software SVM_V251 is used for training. Four ranges of values have been selected for soft margin parameter © in Eq. 3, namely 10-25, 25-50, 50-75

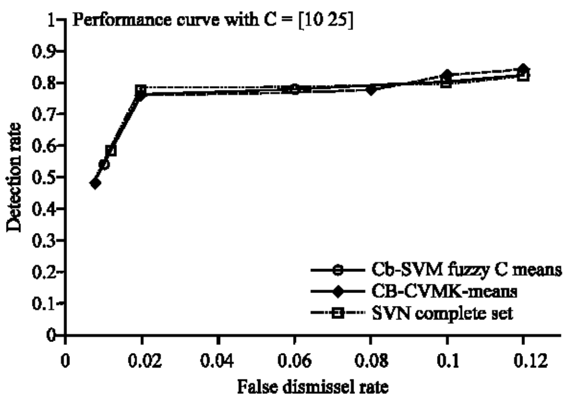

Fig. 3: Performance characteristics C:[10 25]

and 75-100.The performance curve is drawn by increasing kernel parameter . in Eq. 5 from 10-2500 for each case of C. Figure 3 presents the performance characteristics curves of the Gaussian RBF when the range of C = [10-25], on the test set after it has been trained with complete training data set with and without data clustering. On X-axis we have taken the false dismissal rate, which is the ratio of test points incorrectly classified as stable. The ratios of test points have been correctly classified as stable called detection rate has been taken on the Y-axis. The false dismissal case occur when the neural network classifies an unstable case as stable and false alarm case occur when it conclude a stable case as unstable. Here SVM and CB-SVM both K-Means and Fuzzy C-Means provide closer results.

Table 1 shows the performance of the SVM and CB-SVMs, in terms of detection rate, False dismissal rate, Training time, False Alarm rate and Error rate for three ranges of C, when . equals 1500 and the number of clusters in CB-SVM equals three. Error rate is the sum of the false dismissal rates and false alarm rates calculated as a percentage of total 150 test patterns. In case I, the K-Means clustering has least training time. The detection rate is same for the three networks, whereas the false dismissal rate of Fuzzy C-Means clustering SVM is smaller than the other and it has greater false alarm rate. Hence in a performance curve, if we want to reduce the false dismissal rate, the points move to the left side resulting in larger false alarm rate. In case II, the three networks produce the same results. In case III, again the Fuzzy C-Means clustering SVM has least training time and false dismissal rate. This clearly shows that the data Clustering based SVM maintains the same degree of classification accuracy as the SVM trained with complete data set with minimum training time.

In Fig. 4 the performance characteristics with C ranges from 75-100 is depicted. Here Fuzzy C-Means data clustering shows best test results. Out of the four ranges of C, this range provides higher detection rate of 0.94 when . has a value of 2500.

Table 1: Network performance on the test set

	Case I:C=[10 -25]			Case II:C=[50 -75]			Case III:C [75 -100]		
Performance parameter	SVM	CB-SVM (K-M)	CB-SVM (F-M)	SVM	CB-SVM (K-M)	CB-SVM (F-M)	SVM	CB-SVM (K-M)	CB-SVM (F-M)
No of I/P features	22	22	22	22	22	22	22	22	22
Training data patterns	450	450	450	450	450	450	450	450	450
Training time(sec)	0.094	0.047	0.062	0.047	0.047	0.047	0.047	0.031	0.031
Norm of Sep. hyper plane	21.11	21.1307	21.1699	21.11	21.1307	21.1699	21.11	21.1307	21.1699
No of support vectors	445	444	444	445	444	444	445	444	444
False Dismissal Rate(FDR)	0.06	0.08	0.04	0.09	0.09	0.09	0.04	0.06	0.01
Detection Rate (DR)	0.78	0.78	0.78	0.8	0.8	0.8	0.81	0.82	0.84
False alarm rate	0.16	0.14	0.18	0.1	0.1	0.1	0.15	0.12	0.15
Error rate	0.22	0.22	0.22	0.19	0.19	0.19	0.19	0.18	0.16

Table 2: Influence of clustering

	Fuzzy C-means clustering					K-means clustering				
No of clusters	Training time (sec)	No. of SVs	Norm of Sep. Hr. plane	FDR	DR	Training time (sec)	No. of SVs	Norm of Sep. Hr. Plane	FDR	DR
3	0.031	444	21.1699	0.01	0.84	0.031	444	21.1307	0.06	0.82
6	0.028	444	21.1336	0.008	0.86	0.030	446	21.1477	0.05	0.83
10	0.031	446	21.1544	0.008	0.86	0.030	445	21.1094	0.05	0.83
15	0.032	445	21.1144	0.008	0.86	0.031	444	21.1103	0.05	0.83

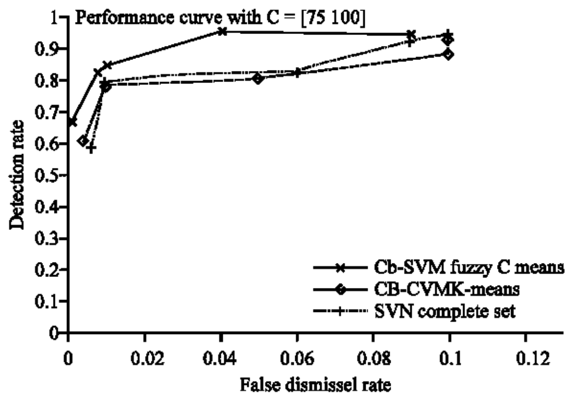

Fig 4: Performance characteristics with C: [75-100]

In order to analyze the change in network behaviour due to the change in the number of clusters, the CB-SVMs are trained with datasets having increased number of clusters. Table 2 renders the network performance with C ranges from 75-100 and . at 1500.

APPENDIX

Table A-1: Generator data of New England 39Bus System

Gen. No	Ra pu	$X_{d'}$ pu	H
1	0.0	0.00647	30.3
2	0.0	0.0060	500
3	0.0	0.0531	35.8
4	0.0	0.0660	26
5	0.0	0.0436	28.6
6	0.0	0.05	34.8
7	0.0	0.049	26.4
8	0.0	0.057	24.3
9	0.0	0.057	34.5
10	0.0	0.004	42

CONCLUSION

Main requirements for online TSA methodologies are fast evaluation time and accuracy of results. In a real power system with billions of data records even multiple scans of the entire data are too expensive to perform. In such instances CB-SVM is best suitable for TSA application. It provides a different strategy to tackle the curse of dimensionality, regarding computational effort, because of low training time and improved classification compared to SVM connected with feature extraction. Thus CB-SVM paves the way to online Transient Stability Assessment and Corrective control actions.

REFERENCES

Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery, 2: 121-167.

Fung, G. and O.L. Mangasarian, 2001. Proximal Support Vector Machine Classifiers. In: Proc. 7th Int. Conf. Knowledrge Discovery and Data Mining, Sanfrancisco, CA.

Jain, A.K., M.U. Murty and P.J. Flynn, 1999. Data Clustering- A Review, ACM Computing Survey, Vol. 31.

Moulin, L.S., A.P. Alves da Silva, M.A. El-Sharkawi and R.J. Marks, 2004. Support Vector Machines for Transient Stability Analysis of Large-Scale Power Systems, 19: 818-824,

Padiyar, K.R., 2000. Power System Dynamics-Stability and Control, BS Publications.

Sobajic, D.J. and Y.H. Pao, 1989. Artificial Neural-net based dynamic security assessment for electric power systems, IEEE. Trans. Power Sys., 4: 220-228.

Veronica, S.M., 2002. Introduction to five data clustering algorithms, Integral, Vol. 7.

Yoh-Han Pao and D.J. Sobajic, 1992. Combined Use of supervised and unsupervised learning for Dynamic security assessment, IEEE. Trans. Power Sys., 7: 525-532.

Zhou, Q., J. Davdsonand and A.A. Fauod, 1994. Application of artificial Neural Network in power system security and vulnerability assessment, IEEE. Trans. Power Sys., 9: 525-532.