Data Mining and Soft Computing using Support Vector Machine : A Survey

Subhankar Das
Student of M.Tech, CST
Department of Computer Science & Engineering
Dr. B.C. Roy Engineering College, Durgapur, India

Sanjib Saha
Assistant Professor
Department of Computer Science & Engineering
Dr. B.C. Roy Engineering College, Durgapur, India

ABSTRACT

In this paper, the basic concepts and survey of the available literature on Support Vector Machines (SVM) in data mining and soft computing research area is provided. While at the time of survey several new methods were found related to SVM like as Support Vector Representation and Discrimination Machine (SVRDM), Recursive SVM (RSVM), On-line Independent SVM (OISVM), Pruning SVM, Fast Nearest Neighbor Condensation classifier (FCNN-SVM), Improved SV Clustering (iSVC), Cost-sensitive SVM (2v-SVM), 2C-SVM, Profile SVM (PSVM), Twin SVM (TWSVM), Twin Bounded SVM (TBSVM), Parametric-margin n-SVM (par-n-SVM), Twin Parametric-Margin SVM (TPMSVM), Structural Twin SVM (S-TWSVM), Hierarchical Linear SVM (H-LSVM), Bio-SVM, FuzzySVM-CIL, Kernel Fuzzy C-Means clustering-based Fuzzy SVM (KFCM-FSVM), Multi-Class Instance Selection (MCIS). After studied these methods a comparative and analytical survey upon those methods are presented here. Also a large future scope is available on several techniques and they are discussed in this paper.

Keywords

Data Mining, Soft Computing, SVM, Maximum Margin, Soft-Margin, Kernel Trick.

1. INTRODUCTION

Support vector machine (SVM) is a popular machine learning and data mining classification, regression method [1]. Margin maximization and systematic nonlinear classification via kernel tricks are developed using SVM. There are several features of SVM.

Linearly Separable Data: Here main goal is to find the hyperplane also called decision boundary that linearly separate the binary classification.

Here boundary equation $w^Tx + b = 0$ where w and b can get from training data. For any two-class problem, when the patterns from the two classes are linearly separable then $w^Tx + b > 0$ where pattern x from the positive class and $w^Tx + b < 0$ where pattern x from the negative class. The reason behind labelling scheme is that it lets condense the formulation for the decision function to $f(x) = sign(w^T x + b)$ since $f(x) = +1$ for all x above the boundary, and $f(x) = −1$ for all x below the boundary.
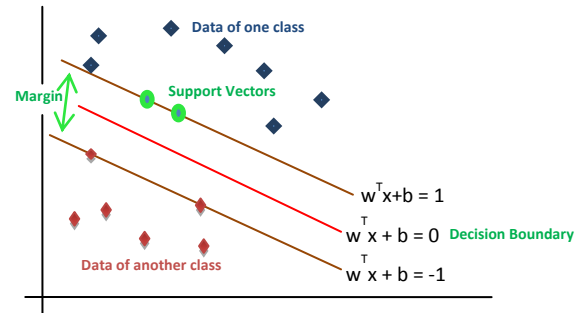
Figure 1: Linearly Separable Hyperplane

Maximum Margin: Positive class linearly separable data can be characterized by a hyperplane $w^Tx + b = 1$ and the negative class by a parallel plane $w^Tx + b = −1$ so that training patterns, near the margin, from the respective classes fall on these planes. These planes are called the support planes and points on the support planes are called support vector. The decision boundary or the separating plane is characterized by $w^Tx + b = 0$.

The distance between newly added boundaries can be found as follows- let x3 be any point on the "minus" hyperplane and let x1 be the closest point to x3. Here two lines are parallel, and thus share their parameters w; b. Pick an arbitrary point x1 to lie on line $w^Tx + b = +1$. Then, the closest point on line $w^Tx + b = −1$ to x3 is the point x1 = x3 + λw (because closest point will always lie on the perpendicular; recall that the vector w is perpendicular to both lines).Using this formulation, λw will be the line segment connecting x1 and x3, and thus, λ‖w‖, the distance between x1 and x3, is the shortest distance between the two lines/boundaries. Solving for λ:

$=>w^T x1 + b = 1$ where x1 = x3 + λw

$=>w^T (x3 + λw) + b = 1$

$=>w^T x3 + b + λw^T w = 1$ where $w^T x3 + b = −1$

$=> −1 + λw^Tw = 1$

$=>λw^T w = 2$

$=> λ = 2 / w^Tw = 2 / ‖w‖^2$

And so, the distance λ‖w‖ is $2 / ‖w‖^2 * ‖w‖ = 2 / ‖w‖$

$= 2 / √w^T w$.

It's intuitive that to maximize the distance $2/√w^Tw$ between the two boundaries demarcating the classes. Similarly, minimizing $√w^Tw/2$, which is in turn equivalent to minimizing $w^Tw/2$ (since square root is a monotonic function). This quadratic programming problem is expressed as: $min_{w,b}(w^Tw/2)$.
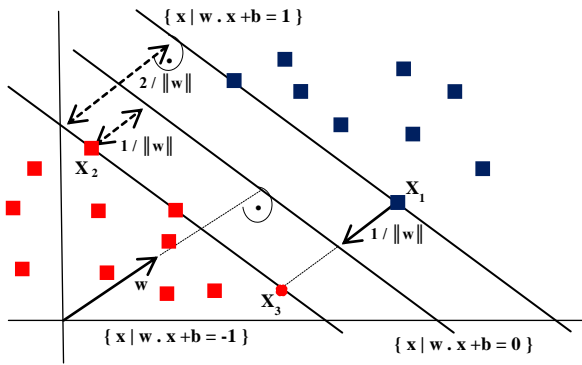
Figure 2: Optimal Hyperplane

Soft-margin Extension: Here the case is considered where data is not perfectly linearly separable. For instance, it cannot be guaranteed that all the data points are correctly labeled, so allow some data points of one class to appear on the other side of the boundary. Slack variables introduced- an $\varepsilon_i \geq 0$ for each x_i. The quadratic programming problem becomes: $\min_{w,b,\varepsilon}(w^T w/2) + C \sum_i \varepsilon_i$.

Nonlinear Decision Boundary: It is possible that the two classes are not linearly separable. In that case, the decision boundary can be non-linear. In general, a non-linear function can be represented by a linear function. Mapping the data vectors, x_i, into a higher-dimension (even infinite) feature space may make them linearly separable in that space (whereas they may not be linearly separable in the original space). The formulation of the quadratic programming problem is as above, but with all x_i replaced with $\mu(x_i)$, where μ provides the higher-dimensional mapping. So the standard SVM formulation will be: $\min_{w,b,\varepsilon}(w^T w/2) + C \sum_i \varepsilon_i$.

Linearly Nonseparable: It is possible that the classes may not be linearly separable in some cases. To solve this, convert the d-dimensional patterns into D-dimensional patterns where D>d.
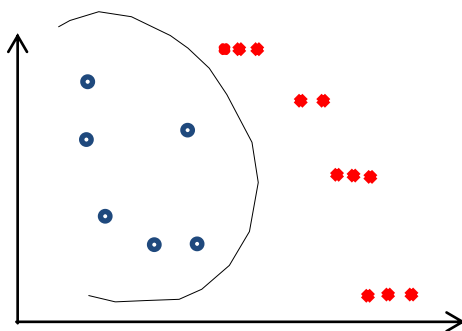
Figure 3: Linearly Non-separable Hyperplane

Kernel Trick: A higher-dimension space is required for working (and potentially even an infinite-dimensional space), calculating $\mu(x_i)^T \mu(x_j)$ may be intractable. However, it turns out that there are special kernel functions that operate on the lower dimension vectors x_i and x_j to produce a value equivalent to the dot product of the higher-dimensional vectors.

Support vector machines are widely used in fraud credit card detection, finger print matching, handwritten digit recognition etc. SVMs have also been successfully applied to an increasingly wide variety of biological applications. A common biomedical application of support vector machines is the automatic classification of microarray gene expression profiles.

2. DATA MINING AND SVM

Data Mining refers to the process or method that extracts or mines interesting knowledge or patterns from large amounts of data [2]. Knowledge Discovery from Data or KDD is referred as synonym of data mining. Data Mining can be categorized in following ways: Association, Classification, Clustering, Regression, and Outlier Detection. **Association** gives the correlation between frequent items or patterns and generates certain rules based on support, confidence value of frequent items. **Classification**, classifies test data applying classification model after training of the data which have pre-defined class. Therefore, it is called supervised learning. **Clustering** groups similar types of data or patterns in clusters. In clustering, data have no pre-defined class. Hence it is called unsupervised learning. **Regression** is a statistical method that predicts the value of a dependent variable from one or more independent variables where the variables are numeric. **Outlier** is that the data objects which deviates significantly or inconsistent from the normal objects as if it were generated by a different mechanism. To detect outliers, the behavior of individual objects and groups of objects should be known.

In this paper the review and survey of SVMs in the fields of data mining are provided. After surveyed on various recent or past research papers on SVMs, here a list of summary of those papers are given and from reference the reader can easily get the detail about those paper on which he/she will find interest. From available future work, the reader can easily extend that particular papers method or idea for better results in future.

A new hierarchical design method, weighted support vector (WSV) k-means clustering is used to design a binary hierarchical classification structure [3]. It automatically selects separate classes at each node in hierarchy. An SVRDM (Support Vector Representation and Discrimination Machine) classifier is used at each node in hierarchy. It offers generalization and good rejection of unseen false objects. The new method and SVRDM binary classifiers both operate in high dimensional space and can find the cause of best rejection by Gaussian kernel. This hierarchical classifier gives better perfect classification P_C and false alarm P_{FA} than standard one-vs-rest classifiers for its simplicity and it is more efficient in computation.

A recursive SVM (RSVM) with several orthogonal directions best separates data with maximum margin [4]. Theoretical analysis proved that completely orthogonal basis can be derived in feature subspace spanned by training samples. Through RSVM a new dimensionality reduction technique is achieved based on multilevel maximum margin components. Compared with linear discriminant analysis (LDA) and regular SVM this method has no singularity problems.

Modification of v-support vector machines (v-SVM) for regression and classification is described here; parametric insensitive/margin model usage with an arbitrary shape is demonstrated [5]. A new regression algorithm par-v-SVR automatically adjusts a flexible parametric-insensitive zone of arbitrary shape and minimal radius to include given data by using a parametric-insensitive loss function. This algorithm is useful in case of noise is heteroscedastic. Based on this algorithm a new algorithm can be developed named par-v-SVC for classification and it uses a parametric-margin model of arbitrary space.

A convex energy-based framework is introduced to jointly perform feature selection and SVM parameter learning for linear and non-linear kernels by using a convex method [6]. The proposed framework is related to L1-SVMs and provided theoretical justification for its use as a feature selection method. This method gives better result in classification when it produces SVM classifiers which used sparse sets of features and support vectors.

On-line independent support vector machines (OISVMs) approximately converges to standard SVM solution each time new observations are added; approximation is controlled via a user-defined parameter [7]. It takes less time and less space and gives better accuracy. Here the optimization problem is solved by an incremental algorithm; it gives benefit as it has small size of basic set. It gives better results in case of place recognition in an indoor environment and human grasping classification. OISVM able to perform continuous learning on data collected on a span of time up to several months but at this time it is not possible to use them with a truly never-ending stream of data, since all training sample must anyway be retained. To achieve this problem is another future scope.

It is a practical and effective method of pruning SVM classifiers that is developed systematically [8]. The pruning process circumvents the problem of explicitly discerning SVs in the high-dimensional feature space and does not involve local minima. This pruning method is able to extract a more compact SVM architecture without affecting the performance of the trained SVM. The pruning method can be used for incremental online training.

FCNN-SVM classifier combines the support vector machine (SVM) approach and the fast nearest neighbor condensation classification rule (FCNN) for making SVMs practical usage on large amount of data [9]. FCNN works faster on very large and multidimensional data sets and the requirement of SVs is more than halved compared to SVM. It is used when fast response time is a basic requirement. This classifier reduces the training time very efficiently and the loss of accuracy is very minimal respect to standard SVM.

In this algorithm a multidimensional feasible space for the optimality condition is used to solve the tracking problem [10]. It can handle datasets in an efficient way. Data sets are characterized by the presence of linearly dependent points, duplicate points, or nearly duplicate points. It gives better result compared to SVM-path algorithm in case of dealing with data sets.

In this paper it is proved that Improved Support Vector Clustering (iSVC) overcomes two bottlenecks well [11]. And compared to some common clustering methods it gives better efficiency on performance and it opens a broad way of applications for SVC. iSVC consists of a reduction strategy and a new labeling approach which is based on geometric property of feature space. Experiments show that a good behavior of iSVC overcomes two bottlenecks and performs fine on clustering applications.

LSVM constructs multiple linear SVMs instead of building sophisticated global model from training data, each linear SVM is designed to accurately classify a given test example [12]. For reducing high computational cost an efficient LSVM is introduced, named Profile SVM (PSVM). A framework is proposed for PSVM that performs better over both KNN and nonlinear SVM. PSVM minimizes the computational time by extracting a small number of clusters using a supervised algorithm called MagKmeans.

Here this is an extension of incremental decremental algorithm which efficiently works for simultaneous update of multiple data points in case of online learning [13]. This is specially useful for online SVM learning and in that case we have to remove old data points and add new data points in a short time. In this multiple update algorithm the number of break points is much smaller than that of single update algorithm. In many situations it is also faster than LIBSVM (Library for Support Vector Machine) and gives more accurate result.

Minimax and Neyman-Pearson classification can be optimized in a straightforward way using a cost-sensitive Support Vector Machines (2v-SVM) and smoothed the cross-validation error estimates with compare to usual cost-sensitive Support Vector Machines (2C-SVM) [14]. Simple smoothing techniques lead to significantly improved error estimates, which translate into better parameter selection and a dramatic improvement in performance. This smoothing approach also performs a weighted local averaging to reduce outlying estimates.

When the likelihoods are unknown and training dataset is given but no equivalent result to the Neyman–Pearson lemma to take a unique decision for K test samples, there are two alternatives to solve this problem [15]. First, the consensus decision takes the posterior probability estimates to predict a single label for K test samples, and second, build a new classifier that directly computes the label for K test samples. Second approach need to define an extended input space training set and incorporate the known symmetries in the classifier and gives more accurate results, as it only requires an accurate classification boundary, while the former needs an accurate posterior probability estimate for the whole input space.

Here introduced a time-adaptive support vector machine (TA-SVM), is a new method for generating adaptive classifiers, capable of learning concepts that change with time [16]. Compared to other several problems including well-known STAGGER dataset and the real-world electricity pricing domain, shows the good performance on TA-SVM in all tested situations. At present studying the application of TA-SVM in real problems in slowly drifting systems is particularly for fault prediction in critical mechanical systems.

This TPMSVM, in the spirit of the twin support vector machine (TWSVM), determines indirectly the separating hyperplane through a pair of nonparallel parametric-margin hyperplanes solved by two smaller sized support vector machine (SVM)-type problems [17]. This TPMSVM is suitable for many cases, especially when the data has heteroscedastic error structure. The experimental results on several artificial and benchmark datasets proved that the TPMSVM obtains fast learning speed, and also shows good generalization. Recently presented a sparse TSVR in primal space based on a simple back-fitting strategy. The same concept can be applied in the TPMSVM to improve the sparsity of its parametric margin hyperplanes.

Tuning of the regularization parameter C is a well-known process in the implementation of a support vector machine (SVM) classifier [18]. This tuning process determines the global minima C values of common validation functions for SVM classifier over a validation set or cross-validation set. This was possible because the SVM solution path for a range of C values could be computed. When the solution of the SVM solution path is available, the approach is generally very fast.

In this paper, unify classifiers into a common framework from the concept of "structural granularity" and the formulation for optimization problems are described [19]. By exploiting the quadratic programming (QP) and second-order cone programming (SOCP) methods and deriving a novel large margin classifier a new classifier is produced namely structural regularized support vector machine (SRSVM). A novel algorithm SRSVM was derived from the cluster granularity in the frame work, it captures the data structural information within individual classes by some clustering strategies. The corresponding optimization problem follows the same QP formulation as SVM. The experimental results proved that SRSVM gives better results compared to the state-of-the-art algorithms in the framework. There are several future scopes are present in the paper:

a) **Additional generalization:** The combinations of different granularities, such as cluster granularity and point granularity, may lead to a large family of new algorithms.

b) **Large-scale problem:** How to develop a fast algorithm for SRSVM to solve large-scale problems is another interesting topic for future study.

c) **Prior knowledge:** Suitable classifier selection between the two algorithms actually depends on more prior knowledge about the data. To develop more effective methods to reveal more prior knowledge hidden in the data to guide classifier design.

A new matrix pattern based MCSSVM (MCSVMmatrix) is presented through minimum within-class scatter support vector machines (MCSVM) [20]. This is extended by introducing Mercer's kernels for solving the problem of nonlinear decision boundaries, which presents a significant matrix pattern based nonlinear support vector machines: Ker-MCSVMmatrix. This approach is more useful for solving the singular problem, when dealing with small sample size problem it reduces the time/space complexity when within-class scatter matrix, its invertible matrix and weight vector are calculated. So the classification accuracy is improved and small size high-dimensional datasets can be handled easily. In this method though the iteration speed is faster but classification accuracy is not enhanced significantly, so there is a future work to find out a better optimization method to enhance the classification accuracy.

Here a new improved version of twin support vector machine (TWSVM) is introduced namely twin bounded support vector machines (TBSVM) [21]. The main advantage of this TBSVM over TWSVM is that by introducing the regularization term the structural risk minimization principle is implemented. This modification can improve the classification performance because it implements the marrow of statistical learning theory. Here the successive over relaxation (SOR) technique is used to solve the optimization problem for speeding up the training procedure. By using SOR technique the drawback of TWSVM can be overcome. Compared to TWSVM and SVC this TBSVM method is faster and also shows better generalization.

Here it exploits for the privacy violation problem of SVM classifier and propose an approach to post process the SVM classifier to transform it to a privacy-preserving classifier which does not disclose the private content of support vectors [22]. The post processed SVM classifier without exposing the private content of training data is called Privacy-Preserving SVM Classifier (PPSVC). It is proved that PPSVC is robust against adversarial attacks and classification accuracy is same as original SVM classifier. The resulted PPSVC can be publicly released for its protecting nature of sensitive content of support vectors. For future work the PPSVC can be applied to high-dimensional data, and exploiting on other common kernel functions like the polynomial kernel.

Background modeling and subtraction is a natural technique for object detection in videos captured by a static camera [23]. A pixel wise generative background model is obtained for each feature efficiently and effectively by Kernel Density Approximation (KDA) where generative and discriminative techniques are combined for classification. Background subtraction is performed in a discriminative manner using a SVM over background likelihood vectors for a set of features and here a new algorithm is proposed for this. KDA represents a probability density function of the background for RGB, gradient, and Haar-like features in each pixel, where 1D independent density functions are used for simplicity. This algorithm gives better performance than other density-based techniques such as Gaussian Mixture Model (GMM) and Kernel Density Estimation (KDE).

In applications such as credit card fraud detection, the classifier needs to predict an event in 10ms at most [24]. In this case the speed of prediction constraints heavily outweighs the training costs. The new classification method, called Hierarchical Linear Support Vector Machine (H-LSVM), provides a very simple and efficient model in training but mainly in prediction for large-scale datasets. Only a few hyperplanes need to be evaluated in the prediction step and no kernel computation is required and the tree structure makes parallelization possible. The H-LSVM algorithm improves the classification accuracy of linear SVMs while working on medium and large datasets. Compared with existing methods based on the construction of a decision tree with linear SVMs as splitting criteria, the H-LSVM performs better in terms of classification accuracy while maintaining a classification complexity of the same order of magnitude. The H-LSVM method is an attempt to solve the problem of applying SVM technology to industrial settings with high loads in real-time classification. In online industrial environments for taking quick decision the non-linear SVMs are not suitable so H-LSVM makes a bridge for this gap because its simplicity.

The new method multitask multiclass support vector machines (M2SVMs) based on the minimization of regularization functional is introduced [25]. Cast multitask multiclass problems into a constrained optimization problem with a quadratic objective function. This approach can learn in two different ways: label-compatible and label-incompatible multitask learning. This method easily generalize the linear multitask learning method to the non-linear case using kernels. It derived a dual optimization problem which is optimized to give solutions. Then it extends the method to non-linear multi-task learning by the kernel trick. Compared to other two popular multitask learning methods which are based on one2one and one2all strategies shows that this method performs better than those. How to extend this model to domain adaption and transfer learning is a scope on it.

To overcome problems such as SVM not preferred for applications requiring high classification speed or limited computational resources present in system, it has been introduced a primal SVM method with the following properties: [26]

a) Solves for the SVM representation without need to invoke representer theorem.

b) Forward and backward selections are combined to approach the final globally optimal solution.

c) A criterion is introduced for identification of support vectors leading to a much reduced support vector set.

Forward and backward selections are combined to identify the support vector set for the solution SVM. A criterion (threshold) for support vector selection is defined that based on the violation of a point. The proposed algorithm is tested and compared with LIBSVM, SVMPerf and SpSVM (Sparse SVM) on three public benchmark problems and a human activity recognition application and shows that this algorithm gives better results than those algorithms.

Additive kernels include widely used kernels for histogram-based image comparison like intersection and chi-squared kernels [27]. Additive kernel SVMs can offer improvements in accuracy over linear SVMs on a wide variety of tasks while having the same runtime, making them practical for large-scale recognition or real-time detection tasks. The proposed technique can be applied to settings where evaluation of weighted additive kernels is required, which include kernelized versions of PCA, LDA, regression, k-means, and also for speeding up the inner loop of SVM classifier training algorithms. An additive kernel is strictly more general than linear kernels and often provides significant improvements. It is widely used in computer vision and this technique has found widespread applications in many classification or detection tasks.

For optimizing the support vector machines performance a general procedure is suggested to produce nonparametric and efficient kernels [28]. This is achieved by finding an empirical and theoretical connection between positive semi-definite matrices and certain metric space properties. This paper suggested a new direction to generate nonparametric and efficient kernels. By implementing several kernels using a large variety of binary real-world and artificial datasets has demonstrated that the kernels generated using the new framework provide a classification rate superior to that of most of the well-known kernels defined thus far. Also this kernel need less training time. In future some open problems can be solved such as proving the generalization of proposition 1 to n×n matrices, determining the necessary and/or sufficient conditions for the decreasing sequences and generalizing the suggested framework to matrices with an arbitrary diagonal.

A new Structural Twin Support Vector Machine (S-TWSVM) is introduced in this paper [29]. Unlike existing methods based on structural information, S-TWSVM uses two hyperplanes to decide the category of new data, of which each model only considers one class's structural information and closer to the class at the same time far away from the other class. This makes S-TWSVM more efficient and improves the capacity of generalization of the algorithm. Theoretical analysis and all experimental results proved that S-TWSVM can more fully exploit this prior structural information to improve the classification accuracy. The extension of semi-supervised learning and multi-instance classification can be considered as a future work.

Instance selection methods have been proposed and shown significant efficacy to reduce the training complexity of SVM, but more or less trade off the generalization performance [30]. It presents an instance selection method especially for multi-class problems. The purpose of clustering is to improve the efficiency of instance selection. After experiments on variety of datasets proved that the proposed method selects fewer instances than most competitive algorithms and provides highest classification accuracy on most datasets. It also performs superiorly for binary problems. The new instance selection method, Multi- Class Instance Selection (MCIS), selects boundary instances as training data for substantially reducing the scale of a multi-class dataset and effectively speeds up the training of SVM models. For determining parameters of MCIS effectively, a simple skill is presented, by which the possible imbalance of training data caused by o-v-r (one versus rest) style can also be controlled. Like other instance selection algorithms, MCIS mainly accelerates the training process of SVM, though it cans also speedup classification additionally due to the reduction of support vectors caused by it.

3. SOFT COMPUTING AND SVM

Soft Computing is a branch for building intelligent and wiser machines [31]. Purity of thinking machine intelligence, freedom to work, dimensions, complexity and fuzziness handling capability increases day by day. Soft computing can be categorized in following ways: Fuzzy Logic, Evolutionary Computing, Neural Network, and Machine Learning. **Fuzzy Logic** is the application of fuzzy set theory. It gives a two valued form, i.e. true value as 1 and false value as 0. It allows in linguistic form, the set membership values for making concept imprecise like 'slightly', 'quiet' and 'very'. In **Evolutionary Computing** there are different parts like Genetic Algorithm, Ant Colony Optimization and Swarm Optimization. Genetic Algorithms are adaptive heuristic search algorithm and it is based on natural selection and genetics' evolutionary idea. It is better than conventional AI and it proves more robust. Ant Colony Optimization is used for solving computational problems for reducing the findings of good paths in graphs. It has various applications, like in travelling salesman problem it is used to produce near-optimal solutions. Swarm optimization is a computational method for iteratively optimizing any problem to improve any candidate solution based on given quality measure. **Neural Network** concept has been evolved from human biological neuron structure. In soft computing, neural network is a type of network includes many simple processors and each processor should have small amount of local memory. The processors work only on the local data and given inputs which they received from connections. Maximum neural network have some training rule for data. It also has potential for parallelism because computations are maximum independent of each other components. **Machine Learning** is a mixture of computational intelligence and applied statistics. It is mainly three types; these are unsupervised learning, supervised learning and reinforcement learning. Again supervised learning can be of three types as follows, classification learning (pattern recognition), function learning (regression estimation) and preference learning. It has several types of techniques like as statistical learning, artificial neural network (ANN), support vector machine etc. Clustering is an unsupervised learning. In soft computing there are many applications of SVM, here a brief idea is given below of SVM in the area of soft computing.

The new proposed algorithm is an extension of previously proposed hard clustering algorithm and this is based on one class support vector machine (OC-SVM) [32]. A new tool also introduced based on expectation-maximization-type soft clustering algorithm. The algorithm is directly kernelizable for using the WOC-SVM. The proposed algorithm gives best result while performing with several state of-the-art methods,

like spectral clustering algorithm for Ng and Jordan. It also enables to provide robustness respect to initialization. Performance will depend on choice of the kernel for this method. Presently investigating how to adjust kernel to the data in a more automatic way and there is other scope to development of model selection criteria where number the algorithm can easily select number of clusters in data.

It is a biologically feasible support vector machine [33]. It is a bio-SVM model because instant learning of surprising events and off-line tuning of support vector weights trains the system Emotion-based learning, forgetting trivia, sleep and brain oscillations. This model considered as a mechanism for one-shot low-level trainable pattern recognition in the brain. The present model remain consistent with the gross pattern of observed oscillations in the brain, particularly high frequency cortical oscillations combined with the low frequency olfactory "sniff cycle". Components types and functions are available in nature to build up the Bio-SVM classifier. An advantage of Bio-SVM is that learning modifies only the Oscillating Memory (OM) and not feed forward sensory data pathways. Bio-SVM is a SVM architecture based on biologically inspired components and can be more developed by neural classifiers.

The Laplacian support vector machine considered as an extension of the support vector machine to semi-supervised learning [34]. Here the variant of Laplacian support vector machine capable of feature selection based on functional analysis of variance decomposition. Its advantage over LapSVM is that it can select informative features. Also it can achieve better predictive performance than the LapSVM where there are noise features in data. For future the computation can be made more efficient for this method.

Fuzzy SVMs (FSVMs) is a variant of the SVM algorithm, it has been proposed for handling the problem of outliers and noise [35]. In case of FSVMs, training examples are assigned different fuzzy-membership values based on their importance, and these membership values are incorporated into the SVM learning algorithm to make it less sensitive to outliers and noise. FSVMs can suffer from the problem of class imbalance like the normal SVM algorithm. Here an improved FSVM method is proposed, which is called FSVM-CIL, to learn from imbalanced datasets in the presence of outliers and noise. Here assigned Fuzzy-membership values for training samples to handle both the problems of class imbalance and outliers/noise. Also performed a nonparametric statistical test for validating the significance of performance of the proposed method over the existing CIL methods is considered. From the overall results, it can be concluded that the proposed FSVM-CIL method could give significantly better classification results than the existing imbalance learning methods that is applied for normal SVMs, especially in the presence of outliers/noise in the datasets. As future work, the effectiveness should be checked of the FSVM method with other existing imbalance-learning methods, such as re-sampling techniques, for imbalanced dataset learning.

In one of paper proposed a KFCM-clustering-based FSVM algorithm for classification problems with outliers or noises [36]. The contributions of this paper are as follows. Here at first a common misunderstanding of Gaussian-function-based kernel fuzzy clustering in the high-dimensional feature space is corrected. Then the new kernel fuzzy *c*-means clustering-based fuzzy SVM algorithm (KFCM-FSVM) algorithm is proposed for classification problems with outliers or noises. The experiments are performed on six benchmarking datasets and four artificial datasets for testing the generalization

performance of the new KFCM-FSVM algorithm. Then the ultimate results show that the KFCM-FSVM algorithm gives better performance and it is more robust than the fuzzy *c*-means fuzzy SVM (FCM-FSVM) algorithm based on the FCM clustering in the original input space, SVM, and FSVM for classification problems with outliers or noises. The algorithm is also tested on binary classification problems and it can also be easily applied to multi-classification problems. But it is not useful for large-scale classification problems with outliers or noises. There is another interesting topic that would be to solve the problem for FCM's sensitivity to the initial cluster centroids. The future extension of this KFCM-FSVM algorithm would be to handle the nonlinear function estimation problems with outliers or noises.

To improve the comprehensibility of SVMs, a new rule extraction technique is introduced from SVMs via analyzing the distribution of samples [37]. Here defined the consistent region of samples in terms of classification boundary, and forms a consistent region covering of the sample space. Then a covering algorithm is introduced for extracting compact representation of classes. A rule extraction method is proposed from the SVM to find out that what is black box and how it works. By computing the distance between the sample and the nearest support vector within the same class the consistent region is formed. Then a rule set is found based on the consistent region covering reduction. Experiment results on the artificial and real word datasets show that the new method can improve the comprehensibility effectively without foiling the classification performance compared to the original SVM and is comparable to the direct learning methods and other SVM rule extraction techniques.

4. DISCUSSION

In this paper various topics of data mining and soft computing using SVM are presented. A large research area and many future scopes are available in these fields. In several papers many new methods are introduced for solving various types of problems and the new methods are more helpful in such cases. In case of multi-class classification problems a new support vector based design method is introduced for binary hierarchical classifiers and the new classifier is more efficient in computation. It can be more useful in larger class problems and should give better result from any standard method and it can be also applied for recognition of face, fingerprint, handwritten character etc. Another idea especially for face recognition problem is found, named recursive support vector machine to reduce dimensionality. In this way various works on support vector machines was found. In case of localized support vector machine it constructs multiple linear SVMs to accurately classify a given test example. But in this case computational cost is high because for each test example a unique model must be constructed. To overcome this limitation a new framework is introduced, named profile SVM (PSVM). It works better than KNN and nonlinear SVM. In future by extends this method, it will useful for variety of application domains such as climate modeling and dynamic network analysis. There is a new idea for solving non-stationary problems by using a couple support vector machines. This is a time-adaptive support vector machine, it can generate adaptive classifiers and it can change the learning concept with time. It uses a sequence of classifiers and it shows better result compared to other methods. At present it is being used for fault prediction in critical mechanical systems. In future it may be useful for one-class classification and regression problems. For pattern recognition a new method is introduced named TPMSVM (Twin

Parametric-margin Support Vector Machine), that is an improvement over TWSVM (Twin Support Vector Machine). It has more learning speed than parametric-margin n-support vector machine (par-n-SVM) and is suitable for many cases especially for heteroscedastic error structure data types. But this TPMSVM also loses the sparsity and in future this problem can be improved. Here another improvement is found over the TWSVM and the new method is called twin bounded support vector machines (TBSVM). The main advantage is that in TBSVM, by introducing the regularization term the structural risk minimization principle is implemented here. In TBSVM four parameters are used and parameter selection is so typical so there is a future work to make out this particular problem. For classification another new method is found called structural twin support vector machine (S-TWSVM). It uses two hyperplanes to decide the category of new data. For this a new algorithm S-TWSVM is used. It proves that it is more efficient for classification and it should speed up for better result. For credit card fraud detection the classifier needs to predict an event in 10 ms at most. In this situation the speed of the prediction constraints increases the training costs. For this a new classification method Hierarchical Linear Support Vector Machine (H-LSVM) constructs a decision tree in which the node split is obtained as a Linear Support Vector Machine. It is useful for prediction in large-scale datasets. Because a few hyperplanes are needed for prediction and there do not need any kernel computation because the tree structure makes parallelization. This is an approach for applying the SVM technology in industry real-time classification is needed with high load. In online industrial environments for making quick decision this H-LSVM is really helpful for users. For speeding up the support vector machine fast instance selection is required. In this case to reduce the complexity, a new instance selection method is introduced named Multi-Class Instance Selection (MCIS). This is mainly for multi-class problems. For substantially reducing the scale of a multi-class dataset and to speed up the training of SVM models it selects boundary instances as training data. It also gives better results while performing on binary datasets. In future if MCIS is combined with hierarchical classification technique it can improve classification efficiency of SVM and it can be applied on large number of classes having typical practical problems. This method can also be applied to fuzzy SVM for easy selection of instances.

5. CONCLUSION

SVM based design method binary hierarchical classifier is more efficient in computation for multi-class classification problems like face recognition. Recursive SVM (RSVM) for dimensionality reduction method is another idea for face recognition problems. On-line Independent SVM (OISVM) solve optimization problem by an incremental algorithm with less time and less space and gives better accuracy than standard SVM solution. Pruning SVM method can be used for incremental online training. Fast Nearest Neighbor Condensation classifier (FCNN-SVM) works faster on very large and multidimensional datasets. Improved SV Clustering (iSVC) performs fine on clustering applications. Cost-sensitive SVM (2v-SVM) smoothed the cross-validation error better than 2C-SVM. Profile SVM (PSVM) method overcome the high computational cost of localized SVM which constructs multiple linear SVMs to accurately classify a given test example. Time-adaptive couple SVM is a new idea for solving non-stationary problems. For pattern recognition a new method, Twin Parametric-Margin SVM (TPMSVM) that is an improvement over Twin SVM (TWSVM). It has more learning speed than Parametric-margin n-SVM (par-n-SVM).

Another improvement is found over the TWSVM and the new method is called Twin Bounded SVM (TBSVM). The Structural Twin SVM (S-TWSVM) uses two hyperplanes to decide the category of new data. Hierarchical Linear SVM (H-LSVM) constructs a decision tree in which the node split is obtained as a linear SVM. It is useful for prediction in large-scale datasets. Bio-SVM is a emotion-based learning from brain that modifies only the oscillating memory. FuzzySVM-CIL method learns from imbalanced datasets for handling the problem of outliers and noise. Also Kernel Fuzzy C-Means clustering-based Fuzzy SVM (KFCM-FSVM) algorithm is proposed for classification problems with outliers or noises. Multi-Class Instance Selection (MCIS) method is substantially reduce the scale of a multi-class dataset and to speed up the training of SVM models by selecting boundary instances as training data. At present we are trying to find a new method for combining the MCIS and hierarchical classification technique, so that classification efficiency of SVM can be improved on large datasets.

6. REFERENCES

[1] V.N. Vapnik, "The Nature of Statistical Learning Theory", Springer-Verlag, New York, 1995.

[2] Jiawei Han, Micheline Kember, Jian Pei, "Data Mining Concepts and Techniques", 3rd Edition, Morgan Kaufmann, 2012.

[3] Yu-Chiang Frank Wang, David Casasent, "New support vector-based design method for binary hierarchical classifiers for multi-class classification problems" Neural Networks 21 (2008) 502–510, 2008 Special Issue.

[4] Qing Tao, Dejun Chu, Jue Wang, "Recursive Support Vector Machines for Dimensionality Reduction" IEEE Transactions on Neural Networks, Vol. 19, No. 1, January 2008.

[5] Pei-Yi Hao, "New support vector algorithms with parametric insensitive/margin model", Neural Networks 23 (2010) 60-73.

[6] Minh Hoai Nguyen, Fernando de la Torre, "Optimal feature selection for support vector machines", Pattern Recognition 43 (2010) 584–591.

[7] Francesco Orabona, Claudio Castellini, Barbara Caputo, LuoJie, Giulio Sandini, "On-line independent support Vector machines', Pattern Recognition 43 (2010) 1402-1412, ScienceDirect.

[8] Xun Liang, Senior Member, IEEE, "An Effective Method of Pruning Support Vector Machine Classifiers", IEEE Transactions on Neural Networks, Vol. 21, No. 1, January 2010.

[9] Fabrizio Angiulli, Annabella Astorino, "Scaling Up Support Vector Machines Using Nearest Neighbor Condensation", IEEE Transactions on Neural Networks, Vol. 21, No. 2, February 2010.

[10] Chong-Jin Ong, Shiyun Shao, Jianbo Yang, "An Improved Algorithm for the Solution of the Regularization Path of Support Vector Machine", IEEE Transactions on Neural Networks, Vol. 21, No. 3, March 2010.

[11] Ling Ping, Zhou Chun-Guang, Zhou Xu, "Improved support vector clustering", Engineering Applications of Artificial Intelligence 23 (2010) 552–559, ScienceDirect.

[12] Haibin Cheng, Pang-Ning Tan, Member, IEEE, and Rong Jin, Member, IEEE," Efficient Algorithm for Localized Support Vector Machine", "IEEE Transactions on Knowledge and Data Engineering, Vol. 22, No. 4, April 2010.

[13] Masayuki Karasuyama, Student Member, IEEE, and Ichiro Takeuchi, Member, IEEE, "Multiple Incremental Decremental Learning of Support Vector Machines", IEEE Transactions on Neural Networks, Vol. 21, No. 7, July 2010.

[14] Mark A. Davenport, Student Member, IEEE, Richard G. Baraniuk, Fellow, IEEE, and Clayton D. Scott, Member, IEEE, "Tuning Support Vector Machines for Minimax and Neyman-Pearson Classification", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 32, No. 10, October 2010.

[15] Ricardo Santiago-Mozos, Member, IEEE, Fernando Pérez Cruz, Senior Member, IEEE, and Antonio Artés-Rodríguez, Senior Member, IEEE, "Extended Input Space Support Vector Machine", IEEE Transactions on Neural Networks, Vol. 22, No. 1, January 2011.

[16] Guillermo L. Grinblat, Lucas C. Uzal, H. Alejandro Ceccatto, and Pablo M. Granitto, "Solving Nonstationary Classification Problems with Coupled Support Vector Machines", IEEE Transactions on Neural Networks, Vol. 22, No. 1, January 2011.

[17] Xinjun Peng, "TPMSVM: A novel twin parametric-margin support vector machine for pattern recognition", Pattern Recognition 44 (2011) 2678–2692, ScienceDirect.

[18] Jian-Bo Yang and Chong-Jin Ong, "Determination of Global Minima of Some Common Validation Functions in Support Vector Machine", IEEE Transactions on Neural Networks, Vol. 22, No. 4, April 2011.

[19] HuiXue, Songcan Chen, and Qiang Yang, Fellow, IEEE, "Structural Regularized Support Vector Machine: A Framework for Structural Large Margin Classifier", IEEE Transactions on Neural Networks, Vol. 22, No. 4, April 2011.

[20] GaoJuna, Fu-lai Chung, Wang Shitong, "Matrix pattern based minimum within-class scatter support vector machines", Applied Soft Computing 11 (2011) 5602–5610, ScienceDirect.

[21] Yuan-Hai Shao, Chun-Hua Zhang, Xiao-Bo Wang, and Nai-Yang Deng, "Improvements on Twin Support Vector Machines", IEEE Transactions on Neural Networks, Vol. 22, No. 6, June 2011.

[22] Keng-Pei Lin and Ming-Syan Chen, Fellow, IEEE, "On the Design and Analysis of the Privacy-Preserving SVM Classifier", IEEE Transactions on Knowledge and Data Engineering, Vol. 23, No. 11, November 2011.

[23] Bohyung Han, Member, IEEE, and Larry S. Davis, Fellow, IEEE, "Density-Based Multifeature Background Subtraction with Support Vector Machine", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 34, No. 5, May 2012.

[24] Irene Rodriguez-Lujan, Carlos Santa Cruz, Ramon Huerta, "Hierarchical linear support vector machine", Pattern Recognition 45 (2012) 4414–4427, SciVerse ScienceDirect.

[25] You Ji, Shiliang Sun, "Multitask multiclass support vector machines: Model and experiments", Pattern Recognition 46 (2013) 914–924, SciVerse ScienceDirect.

[26] Jian-Xun Peng, Stuart Ferguson, Karen Rafferty, Victoria Stewart, "A sequential algorithm for sparse support vector classifiers", Pattern Recognition 46 (2013) 1195–1208, SciVerse ScienceDirect.

[27] Subhransu Maji, Member, IEEE, Alexander C. Berg, Member, IEEE, and Jitendra Malik, Fellow, IEEE, "Efficient Classification for Additive Kernel SVMs", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 35, No. 1, January 2013.

[28] Essam Al Daoud, Hamza Turabieh, "New empirical nonparametric kernels for support vector machine classification", Applied Soft Computing 13 (2013) 1759-1765, SciVerse ScienceDirect.

[29] Zhiquan Qi, Yingjie Tian, Yong Shi, "Structural twin support vector machine for classification", Knowledge Based Systems 43 (2013) 74–81, SciVerse ScienceDirect.

[30] Jingnian Chen, Caiming Zhang, Xiaoping Xue, Cheng-Lin Liu, "Fast instance selection for speeding up support vector machines", Knowledge-Based Systems 45 (2013) 1–7, SciVerse ScienceDirect.

[31] S. Rajasekaran, G.A. Vijayalaksmi Pai, "Neural Network, Fuzzy Logic, and Genetic Algorithms - Synthesis and Applications", Prentice Hall, 2005.

[32] Manuele Bicegoa, Mario A.T. Figueiredob, "Soft clustering using weighted one-class support vector machines", Pattern Recognition 42 (2009) 27 – 32, ScienceDirect.

[33] Magnus Jandel, "A neural support vector machine", Neural Networks 23 (2010) 607_613, ScienceDirect.

[34] Sangjun Lee, Changyi Park, Ja-Yong Koo, "Feature selection in the Laplacian support vector machine", Computational Statistics and Data Analysis 55 (2011) 567-577, ScienceDirect.

[35] Rukshan Batuwita and Vasile Palade, "FSVM-CIL: Fuzzy Support Vector Machines for Class Imbalance Learning", IEEE Transactions on Fuzzy Systems, Vol. 18, No. 3, June 2010.

[36] Xiaowei Yang, Guangquan Zhang, Jie Lu, Member, IEEE, and Jun Ma, "A Kernel Fuzzy c-Means Clustering-Based Fuzzy Support Vector Machine Algorithm for Classification Problems With Outliers or Noises", IEEE Transactions on Fuzzy Systems, Vol. 19, No. 1, February 2011.

[37] Pengfei Zhu, Qinghua Hu, "Rule extraction from support vector machines based on consistent region covering reduction", Knowledge-Based Systems 42 (2013) 1–8, SciVerse ScienceDirect.

[38] R P Datta, Sanjib Saha, "An Empirical comparison of rule based classification techniques in medical databases", Working paper, Indian Institute of Foreign Trade (IIFT), India, August, 2011.