Volume 4, Issue 12, December 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Survey: Support Vector Machine and Its Deviations in Classification Techniques

Ms. Snehal S. Joshi[1], Mr. Navnath D. Kale[2]
[1] M.E Student, Department of Computer Engineering, PVPIT College, Bavdhan, Pune, Maharashtra, India
[2] Assistant Professor, Department of Computer Engineering, PVPIT College, Bavdhan, Pune, Maharashtra, India

Abstract: This survey paper surveys and support vector machine in classification technology presents the deviation. Support vector machine used in various applications for matching patterns and prediction models. Here's what we support vector machine, Support vector machine model with support vector machine details about survey and used in support vector machine concepts. The linearly separable attacks can be used for classification include Hyper-plane model. The main functional strategy mean that the size of margin and more correctly it is larger than the pattern on classified With regard to the SVM Multiclass problem to resolve this issue we have a comparative analysis have studied classification technique variations. For this purpose, We have different types of SVM classification algorithm is analyzed. Multiclass SVM binary to generate many of the methods have been studied here. Decomposition-based method: to overcome memory limits, variant based technologies: computational complexity, and handle multiclass classification methods based on multiple classes.

Key Words – SVM, Decomposition Algorithm, Variant Based Techniques, Hyper plane

I. INTRODUCTION

Support Vector Machine SVM, SVM, the best machine learning algorithms which was proposed in 1990 and mostly used for pattern recognition, is one of the. Support vector machine method invented the machine learning techniques of supervision. This classification, highly efficient and cost-effective in various applications are used. SVM machine learning is a mistake. In the algorithm that, given a set of training examples, each related to one of the several type as, A model that predicts that the new SVM training algorithm builds a range of example. SVM learning for the general problem, which is aiming at greater statistical capacity. Pattern recognition based on knowledge data either priorities or statistical information in multiple disciplines in the raw data, data separation is a powerful tool that aims to classify from the. A great deal of attention in the last decade SVM drew and actively applied for different domain applications. General classification, regression or SVMs ranking function are used for learning.

This Approach is based on statistical learning theory and structural risk minimization principle .This approach has the aim of determining the location of decision boundaries also known as a hyper plane that produce the optimal separation of classes. The statistical learning theory provides an outline for studying the problem of achieve knowledge, making forecast, making decisions from a set of data.

The support vector machine usually deals with pattern classification that means this algorithm is used mostly for classifying the different types of the patterns. Now, there is different types of the patterns i.e. Linear and non-linear. Linear patterns are patterns that are easily distinguishable or can be easily separated in low dimension, whereas non-linear patterns are patterns that are not easily distinguishable or cannot be easily separated and hence these types of the patterns need to be further manipulated so that they can be easily separated. This survey of SVM includes the Basic Idea of SVM where the formalization and working methods are given. There is also given working model of SVM which shows how Support vectors and margin works.

II. WHAT IS SUPPORT VECTOR MACHINE

Support vector machine, one of the best machine learning algorithms, which was proposed in 1990 and mostly used for pattern recognition. Also image recognition, speech recognition, text classification, face detection and faulty card detection, etc like many paradigm has applied for classification problems. SVM machine learning is a mistake. In the algorithm that, given a set of training examples, each related to one of the several categories as, A model that predicts that the new SVM training algorithm builds a range of example. SVM learning for the general problem, which is aiming at greater statistical capacity.

In statistical learning theory the problem of supervised learning is formulated as follows. We are given a set of training data {(x1, y1)... (xn, yn)} In Rn x R sampled according to unknown probability distribution P(x, y), and a loss function V(y, f(x)) that measures the bugs, for a given x, f(x) is "predicted" instead of the actual value y.

The problem consists in finding a function f that minimizes the expectation of the error on new data i.e., finding a function f that minimizes the expected error:

∫ V(y, f(x)) P(x, y) dx dy^[5]

SVM has attracted a great deal of attention in the last decade and actively applied to various domain applications. SVMs are typically used for learning classification, regression or ranking function. SVM is based on statistical learning theory and structural risk minimization principal and have the aim of determining the location of decision boundaries also known as a hyper plane that produce the optimal separation of classes. Maximizing the margin and thereby creating the largest possible distance between the separating hyper plane and the instances on either side of it has been proven to reduce an upper bound on the expected generalization error.

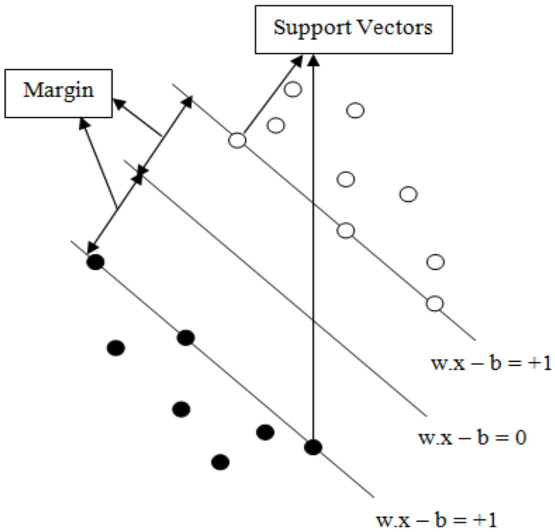

Fig 1-SVM Model

2.1 Basics of SVM

Early machine learning algorithms aimed to learn representations of simple functions. The ability of a hypothesis to correctly classify data not in the training set is known as its generalization. SVM operates better in term of not over generalization when the neural networks might end up over generalizing easily. The support vector machine relates to pattern classification that means the algorithm is used for classifying the different types of patterns. There are two types of patterns i.e. Linear and non-linear. Linear patterns are patterns that are easily distinguishable or can be easily separated in low dimension, whereas non-linear designs are designs that are not easily distinguishable or cannot be easily separated and hence these types of patterns need to be further manipulated so that they can be easily separated.

SVM can also be extended to learn non-linear decision functions by first projecting the input data onto a high-dimensional feature space using kernel functions and formulating a linear classification problem in that feature space. The resulting feature space is much larger than the size of the dataset which are not possible to store in popular computers. Investigation on this issue leads to several decomposition based algorithms. The basic idea of decomposition method is to split the variables into two parts: set of free variables called as working set, which can be updated in each iteration and set of fixed variables, which are fixed at a particular value temporarily. This procedure is repeated until the termination conditions are met.

Basically SVM is based on the construction of the optimal hyper plane, which can be used for classification for linearly separable attacks. The main working strategy at means that if the margin size is larger than it more correctly classifies the patterns.

One of the hyper plane is represented by the following equation:

Hyper plane, $aX + bY = C$ (i)

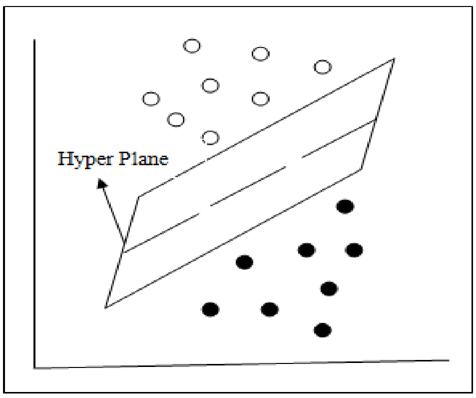

Fig 2- SVM Hyper plane

Selecting different kernel function is an important aspect in the SVM-based classification, commonly used kernel functions include LINEAR, POLY, RBF, and SIGMOID.

Another important parameter in SVM is the parameter C. It is also called a complexity parameter and is the sum of the distances of all points which are on the wrong side of the hyper plane. Basically, the complexity parameter is the amount of error that can be ignored during the classification process. But the value of classification process cannot be either too greater or too small. If the value of complexity parameter is too greater then the performance of classification is low and vice versa.

The main principle of support vector machine is that given a set of independent and identically distributed training sample {(xi, yi)}N i=1, where x ϵ Rd and yi ϵ {−1,1} ,denote the input and output of the classification. The goal is to find a hyper plane wT.x + b = 0, which separate the two different samples accurately. Therefore, the problem of solving optimal classification now translates into solving quadratic programming problems. Where we have to maximize the weight of the margin.
It is expressed as:

$$\text{Min } \Phi (w) = ½ \| w \| 2 = ½ (w, w),$$

Such that: yi (w. xi + b) >= 1 (iii)

2.2 Concepts used in SVM

Concepts of SVM on which SVM is identified are given as bellows:

- The Separating hyper plane.
- The maximum margin hyper plane.
- Soft margin.
- The Kernel function

These Concepts are explained for classification of the given set of patterns by constructing an optimal hyper plane. For any kind of patterns, human beings are considered to be an ultimate judge, who can easily distinguish the different pattern given to them, but for a computer system it is very difficult to distinguish and represent it. In the fig 2.3(a), there are two different kinds of patterns and our job is to classify these two patterns. In this case, it is very easy to classify visually with our naked eye as it can be visually segmented. But, in order to represent these patterns to belong to two different classes, a line can be drawn that separates this pattern.

The fig 2.3(b) shows representation for the classification of two different patterns using a single line, provided that the patterns are presented in two dimensional space. The fig. 3(c) shows the similar type of two different patterns, but in one dimensional space. So, in order to separate these patterns, given in one dimension, a point can be used to separate it. When the similar types of patterns that are presented in fig 3(b) is represented three dimensional space, then a plane can be used to represent a line for the classification of patterns into two different categories as shown in the fig 3(d). The plane that separates these two different types of pattern represented in 3-D space is known as a separating hyper plane that separates patterns.

Similarly, for separating the above mentioned patterns there may exist many such planes as shown in the fig 3(e) that separates the patterns mentioned above. The next task is to select the plane from the set of planes whose margin is maximized. The plane with the maximum margin i.e. perpendicular distance from the marginal line is known as optimal hyper plane or maximum margin hyper plane as shown in fig 3(f). The patterns that lie on the edges of the plane are called support vectors.

During the classification and representation of patterns, there may exist some errors in the representation, as shown in the fig 3(g), such types of errors is called soft margin. During classification of such type of patterns representation, the error can be ignored to some threshold value.

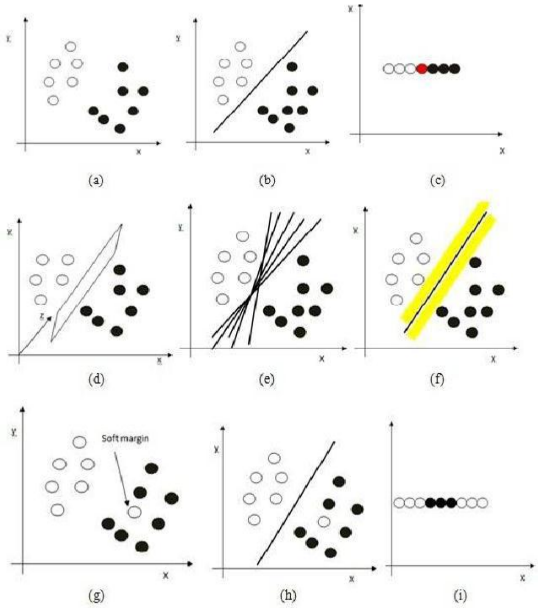

Fig 3- Classification Concept using SVM

The fig 2.3(h) shows the classification of pattern into different categories with soft margin. In other words, it is also called the cost factor or the complexity parameter.

The patterns that are discussed above are all linearly separable patterns that can be easily separated using line or plane. There may also exist non-linear separable patterns that are difficult to classify. The patterns that are discussed above are all linearly separable patterns that can be easily separated using line or plane. There may also exist non-linear separable patterns that are difficult to classify. For such type of patterns to classify, the original data's are mapped to a higher dimensional space using a function called kernel function.

The fig 3(i) shows the representation of pattern that is not linearly separable using a single line or plane. So, in order to classify such types of patterns, the original data's are mapped to a higher dimensional space using kernel function, for example x2 in this case. The fig 3(j) shows the classification of non-linear pattern after mapping the data to two dimensional spaces. The fig 3(k) shows the classification of different non- linear pattern using polynomial function for mapping into higher dimensional spaces. Since it is not easy to represent the pattern for higher dimension, the fig 3(i) shows approximately, the classification, representation of the same data set or pattern in four dimension spaces.

III. DEVIATIONS IN SVM CLASSIFICATION TECHNIQUES

SVM based classification is attractive, because its efficiency does not directly depend on the dimension of classified entities. Though SVM is the most robust and accurate classification technique, there are several problems. Originally, the SVM was developed for binary classification, and it is not simple to extend it for multi-class classification problem. The basic idea to apply multi classification to SVM is to decompose the multi class problems into several two class problems that can be addressed directly using several SVMs. Investigation on this issue leads to several multiclass based algorithms.

Table 1: Decomposition Algorithm

Sr. No.	**Algorithm**	**Key Idea**	**Advantages**	**Disadvantages**
1	SVMlight	many SVM learning problems have much less support vectors than training examples as well as many support vectors which have αi at the upper bound *C*.	1.lower the training time with computational improvement like caching 2. efficient for large scale problem, especially for those with small support vector and most of their αi at the upper bound C.	
2	SMO (Sequential Minimal Optimization)	Considers working set of size 2 in each iteration	1. Each sub problem can be solved analytically without invoking other solvers, thereby convergence is accelerated 2. SMO and its improved versions are effective for large scale SVM training	1. Inefficient due to the use of single threshold value. 2. Very slow for linear SVM
3	Alpha Seeding	Speeding up SVM training by adapting alphas from previous training into appropriate seeds for the next training	1. Training cost is linear in the size of the dataset.	1. Suited particularly for determining penalty coefficient and parameters in kernel using Leave-one-out-cross-validation estimations.
4	LIBSVM	Add b2/2 to the objective function	1. Due to no equality constraint, easier to deal with its dual bound constrained problem 2. comparable with the SVM light in terms of the number of support vectors, the error rate and the optimal value of objective function	Inefficient in performance
5	LASVM	SMO sequential direction search is reorganized	1. Uses less memory 2. Significantly faster than state-of-the-art SVM solver 3. Gracefully handles noisy data 4. Converges to known SVM solution	Make an equal number of process and reprocess iterations which does not guarantee optimal proportion

3.1 Decomposition Based Algorithms

he memory requirement of SVM is grows with the squares of number of training examples. So, the issue is can we scale up the algorithm for large dataset containing thousands and millions of instances .Decomposition based methods break a large optimization problem into a series of smaller Problems, where each problem only involves a couple of carefully chosen variables so that the optimization can be Done efficiently. The following table 3 provides a brief description of some well-known methods implemented to solve scaling problem.

Table- 2: Variant Based Algorithm

Sr. No.	Algorithm	Key Idea	Advantages	Disadvantages
1	LS-SVM (least squared svm)	based on Conjugate gradient scheme	Due to the equality constraints in the formulation, a set of linear equations has to be solved instead of a quadratic programming problem.	suitable for small dataset
2	LSVM (Lagrangian svm)	Used for linear classification	1. Objective function is strongly convex and equality constrain disappear in its dual. 2. Capable of classifying data sets with millions of data in several minutes much faster than SMO and SVM light if the dimension of the input space is small. (Less than 100).	Not able to scale up for very large problems
3	PSVM (Proximal SVM)	Classifies points by assigning them to the closer of two parallel planes that are pushed apart as far as possible.	1. Allows handling very large datasets. 2. Comparable with standard SVM in performance, but fast by several orders of magnitude.	Suited for linear kernel SVM
4	RSVM (Reduced SVM)	Randomly preselect a subset of m examples as support vector candidates	useful for larger problems as well as problems with many support vectors	Remark: Designed for large scale nonlinear kernel SVM.
5	LP-SVM	Changing the metric of margin from 2-norm to 1-norm	Reduce number of dimensions	Convergence rate is similar to simple SVM

3.2 Variant Based Algorithms

Decomposition methods tackle only memory issue by splitting problem into a series of smaller ones, but they are time consuming for large scale problems. Number of methods to reduce the training time have been proposed at the price of accuracy, summarized in table 2.

3.3 Multiclass Based Algorithm:

Originally, SVMs were developed to perform binary classification. However, applications of a binary classification are very limited, especially in remote sensing, land cover classification where most of the classification problems involve more than two classes. A number of methods to generate multiclass SVMs from binary SVMs have been proposed. Here we have compared the different Support Vector Machine techniques. Some of them are decomposition base, some are variant base and some are multi classification base. All the three kinds of techniques are improvements over basic SVM techniques.

Table-3: Multiclass Based SVM algorithm

Sr. No.	Algorithm	Key Idea	Advantages	Disadvantages
1	OVA (one-Against-all)	With the k classes, k binary problems are classified, where each problem discriminates a given class from the other k-1 classes.	Simple, provide comparable performance with other complicated approach when a binary classifier is tuned well.	1. Training complexity is high, as the number of training samples is large 2. Memory requirement is very high during training phase
2	OVO (one-against-one)	Binary classifier requires discriminating between each pair of classes, requiring k (k-1)/2 binary classifiers.	1. Memory required for kernel matrix is smaller	Slower in testing especially when number of classes is big as every test sample has to be presented to large number of classifier

3	DAGSVM	Same idea as OVO, and in recognition phase, the algorithm depends on a rooted Binary DAG to make a decision	Faster Testing and achieving similar recognition rate as OVO	Memory requirement and accuracy is similar to OVA and OVO
4	Error Correcting Codes	Based on idea of error correcting code for neural network	Improve generalization ability	
5	MSVM (Multistage SVM)	Used the support vector Clustering to divide the training data into two parts	Better Generalization capability	Controlling support vector clustering to divide the training dataset into exactly two classes is painful and unfeasible for large datasets
6	HSVM (Hierarchical svm) [23]	Based on clustering classes into a binary tree	Improve performance	1. Knowledge transfer problem 2. How to evaluate stopping criteria for mixed class samples.
7	BTS (Binary tree of SVM)	multiple SVMs arranged in a binary tree structure	Testing time is better than OVO and OVA	Require testing of each trained SVM with all the training samples in order to determine the next step, which significantly increasing the total training time.
8	SVM-BDT	Multiple SVMs arranged in binary structure and based on efficient computation of tree architecture and the high classification accuracy of svm. K-1 svms needed for k class problem.	1. During the recognition phase due to its logarithmic complexity, it is much faster that widely used OVA and OVO methods.	

IV. CONCLUSION

We have reviewed the detail of Support Vector Machine (SVM) and its different Classification Techniques.as the SVM is a supervised learning method. Support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data and identify patterns, used for classification and regression analysis. Different Classification techniques have studied and compared here. Some of them are decomposition base, some are variant base and some are multi classification base. All the three kinds of techniques are improvements over basic SVM techniques. Improvements are proposed by researchers to gain speed, efficiency, space efficiency and ability to handle multiple classes. Every technique holds good in a particular field under particular circumstances. The future work will concentrate on analyzing the performance of different kernel function on different application.

REFERENCES

[1] J. Han and M. Kamber, “Data Mining Concepts and Techniques”, Elevier, 2011.

[2] V. Vapnik and C. Cortes, “Support Vector Network,” Machine Learning, vol. 20, pp. 273-297, 1995.

[3] C. J. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery, vol. 2, 1998.

[4] H. Bhavsar, A. Ganatra, “A Comparative Study of Training Algorithms for Supervised Machine Learning”, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-4, September 2012

[5] G. Wang, “A Survey on Training Algorithms for Support Vector Machine Classifiers”, Fourth International conference on Networked Computing and Advanced Information Management, 2008, IEEE.

[6] G Madzarov, D. Gjorgievikj and I. Chorbev, “ A Multi-class SVM Classifier Utilizing Binary Decision Tree”, Informatica, pp. 233-241, 2009.

[7] M. Aly, “Survey on Multiclass Classification Methods”, November 2005.

[8] V. Vapnik, “Statistical Learning Theory”, Wiley, New York, 1998.

[9] T.Joachims, “Making large-scale support vector machine learning practical”, In Advances in Kernel Methods: Support Vector Machines, 1999.

[10] J.Platt, “Fast training of SVMs using sequ sequential minimal optimization”, In B. Sch¨olkopf, C.Burges and A.Smola (ed.), Advances in Kernel Methods: Support Vector Learning, MIT Press, Cambridge, MA, 1999, 185-208.

