International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 8, August 2012)

SUPPORT VECTOR MACHINE-A Survey

Ashis Pradhan[1]

[1] *Computer Science and Engineering Department, Sikkim Manipal Institute of Technology, Majhitar, East-Sikkim.*

***Abstract*: - Support vector machine (SVM) is one of the most important machine learning algorithms that has been implemented mostly in pattern recognition problem, for e.g. classifying the network traffic and also in image processing for recognition . Lots of research is going on in this technique for the improvement of Qos (quality of service) and in security perspective. The latest works in this field have proved that SVM performs better than other network traffic classifier in terms of generalization of problem. This paper presents a theoretical aspect of SVM, its concepts and its applications overview.**

***Keywords:-* Support Vector Machine (SVM), Machine learning algorithm, Quality of Service, Security perspective, Network Traffic Classification.**

I. INTRODUCTION

Support Vector Machine (SVM), is one of best machine learning algorithms, which was proposed in 1990's and used mostly for pattern recognition. This has also been applied to many pattern classification problems such as image recognition, speech recognition, text categorization, face detection and faulty card detection, etc. Pattern recognition aims to classify data based on either a priori knowledge or statistical information extracted from raw data, which is a powerful tool in data separation in many disciplines. SVM is a supervised type of machine learning. algorithm in which, given a set of training examples, each marked as belonging to one of the many categories, an SVM training algorithm builds a model that predicts the category of the new example. SVM has the greater ability to generalize the problem, which is the goal in statistical learning.

The statistical learning theory provides an outline for studying the problem of gaining knowledge, making predictions, making decisions from a set of data. In statistical learning theory the problem of supervised learning is formulated as follows. We are given a set of training data $\{(x_1, y_1)...(x_n, y_n)\}$ in $R^n \times R$ sampled according to unknown probability distribution P(x, y), and a loss function V(y, f(x)) that measures the error, 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:

$$\int V(y, f(x))\, P(x, y)\, dx\, dy$$ [5].

Early machine learning algorithms aimed to learn representations of simple functions. Hence, the goal of learning was to output a hypothesis that performed the correct classification of the training data and early learning algorithms were designed to find such an accurate fit to the data[6]. The ability of a hypothesis to correctly classify data not in the training set is known as its generalization. SVM performs better in term of not over generalization when the neural networks might end up over generalizing easily [7].

II. SVM MODEL

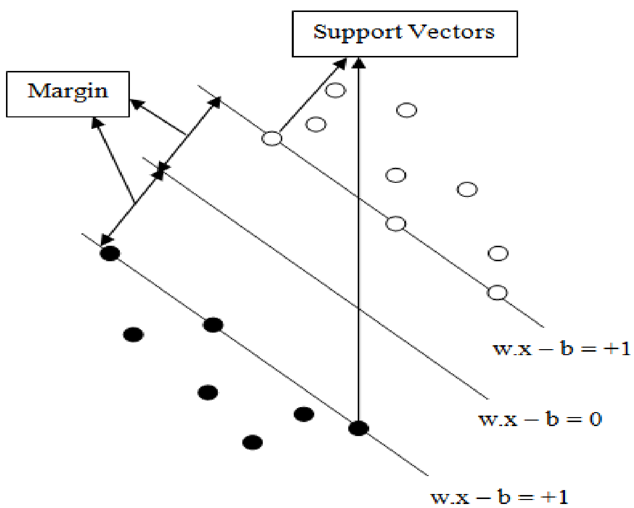

Figure 1: SVM model [8]

The figure 1 is the simple model for representing support vector machine technique. The model consists of two different patterns and the goal of SVM is to separate these two patterns. The model consists of three different lines. The line w.x-b=0 is known as margin of separation or marginal line.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 8, August 2012)

The lines w.x - b = 1 and w.x – b = -1 are the lines on the either side of the line of margin. These three lines together construct the hyper plane that separates the given patterns and the pattern that lies on the edges of the hyper plane is called support vectors. The perpendicular distance between the line of margin and the edges of hyper plane is known as margin. One of the objectives of SVM for accurate classification is to maximize this margin for better classification. The larger the value of margin or the perpendicular distance, the better is the classification process and hence minimizing the occurrence of error.

III. Basic Idea Of SVM

The support vector machine usually deals with pattern classification that means this algorithm is used mostly for classifying the different types of patterns. Now, there is different type 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 patterns are patterns that are not easily distinguishable or cannot be easily separated and hence these type of patterns need to be further manipulated so that they can be easily separated.

Basically, the main idea behind SVM is the construction of an optimal hyper plane, which can be used for classification, for linearly separable patterns. The optimal hyper plane is a hyper plane selected from the set of hyper planes for classifying patterns that maximizes the margin of the hyper plane i.e. the distance from the hyper plane to the nearest point of each patterns. The main objective of SVM is to maximize the margin so that it can correctly classify the given patterns i.e. larger the margin size more correctly it classifies the patterns.

The equation shown below is the hyper plane representation:

$$\text{Hyper plane,} \quad aX + bY = C \qquad (\text{i})$$

The figure 2 shown below is the basic idea of the hyper plane describing how it looks like when two different patterns are separated using a hyper plane, in a three dimension. Basically, this plane comprises of three lines that separates two different in 3-D space, mainly marginal line and two other lines on either side of marginal lines where support vectors are located.

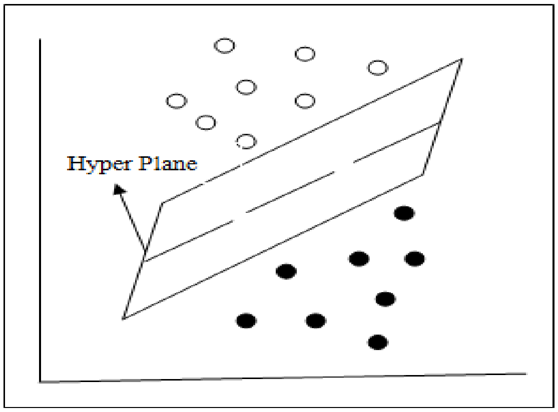

Figure 2: A Hyper Plane [8]

For non-linear separable patterns, the given pattern by mapping it into new space usually a higher dimension space so that in higher dimension space, the pattern becomes linearly separable. The given pattern can be mapped into higher dimension space using kernel function, Φ(x).

$$\text{i.e.} \quad x \longrightarrow \Phi(x)$$

Selecting different kernel function is an important aspect in the SVM-based classification, commonly used kernel functions include LINEAR, POLY, RBF, and SIGMOID. For e.g.: the equation for Poly Kernel function is given as:

$$K(x, y) = \langle x, y\rangle\text{^}p \qquad (\text{ii})$$

Different Kernel functions create different mapping for creating non-linear separation surfaces. 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 large or too small. If the value of complexity parameter is too large 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 $\{(x_i , y_i)\}^N$ i=1, where $x \in R^d$ and $y^i \in \{-1,1\}$,denote the input and output of the classification. The goal is to find a hyper plane $w^T.x + b = 0$, which separate the two different samples accurately.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 8, August 2012)

Therefore, the problem of solving optimal classification now translates into solving quadratic programming problems. It is to seek a partition hyper plane to make the bilateral blank area (2/||w||) maximum, which means we have to maximize the weight of the margin. It is expressed as:

$$\text{Min } \Phi(w) = \tfrac{1}{2} \| w \|^2 = \tfrac{1}{2} (w, w),$$

Such that: $y_i (w. x_i + b) >= 1$ (iii)

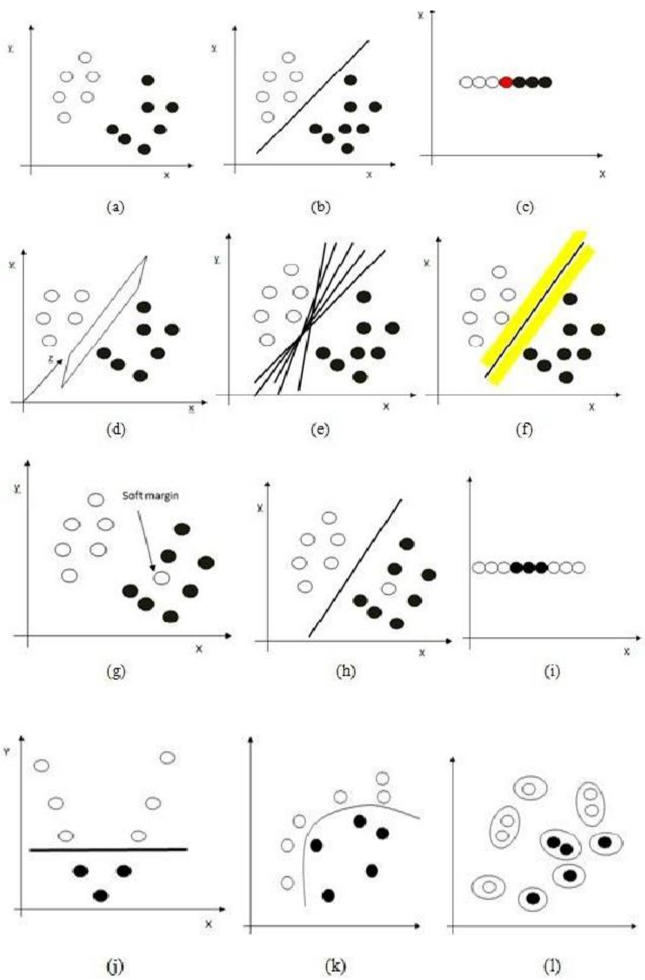

Figure 3: Classification concept using SVM [13]

IV. Concepts Of SVM

The basic concept of SVM can be explained using four points shown below for classification of given set of patterns by constructing an optimal hyper plane.

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

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 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 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 maximum. 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. The fig 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.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 8, August 2012)

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 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 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 same data set or pattern in four dimension spaces.

V. Applications

Since, Support Vector Machine is supervised machine learning algorithm which performs on training basis, so it has mostly been implemented in network areas. For E.g.: classifying the different network application like FTP, HTTP, P2P, etc [4]. The other works of SVM are:-Text classification, Speech recognition, Image clustering for image compression and also Image classification, hand written digit recognition problem, and many other application that requires pattern recognition technique. The SVM can also be implemented in BOTNET detection for isolation of malicious traffic, for improvement in network traffic security. Also some works can be implemented using SVM by filtering network traffic to enhance Qos (Quality of Service). The latest works in this algorithm have proven it that it can be used in recognition of shape and hand gesture in static and also in dynamic environment [12].

VI. Conclusion

The SVM based machine learning technique has been implemented mostly in pattern recognition in networks field as this technique has proven to be efficient than any other most used pattern recognition technique, for e.g. Neural Network, Basiean classification, etc in [2] and [4]. The performance given by the SVM is comparatively higher if it involves large dataset for generalization of problem. The major strength of SVM is that the training of data is relatively easy.

It scales relatively well to high dimensional data and the trade-off between classifier complexity and error can be controlled explicitly. The weakness includes the need of a good kernel-function. It has proven itself best not only in network field but also in image processing for recognition of gesture [11] and there are lots of future scope for this algorithm in image processing for using it in the classification of pattern in multi-dimension.

References

[1] Ruixi Yuan & Zhu Li & Xiaohong Guan & Li Xu "An SVM-based machine learning method for accurate internet traffic classification". Springer Science + Business Media, LLC 2008, Volume 12, Number 2, 149-156, DOI: 10.1007/s10796-008-9131-2.

[2] Zhu Li, Ruixi Yu and Xiaohong Guan "Accurate Classification of the Internet Traffic Based on the SVM Method", Communications, 2007. ICC 2007. IEEE International Conference on, 0.1109/ICC.2007.231, 1373 – 1378.

[3] Li Cong-cong, Guo Ai-ling and Li Dan "Application Research of Support Vector Machine in network Security risk Evaluation", IITAW 2008. International Symposium on, 10.1109/IITA.Workshops.2008.91, 40 – 43.

[4] Andrew W. Moore and Denis Zuevy "Internet Traffic Classification Using Bayesian Analysis Techniques", International conference on Measurement and modeling of computer systems, Banff, Alberta, Canada, Year of Publication: 2005, Pages: 50 – 60.

[5] Theodoros Evgenuiu and Massimilliano Pontil, Statistical Learning Theory: a Primer 1998.

[6] Nello Cristianini and John Shawe-Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods". Cambridge University Press, 2000.

[7] Tom Mitchell, Machine Learning, McGraw-Hill Computer science series, 1997.

[8] Burges, C. J. C. (1998), A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2(2):121–167.

[9] Zhou, Liu, Research on computer network security based on Pattern Recognition, IEEE, 2003.

[10] Tutorial on Support Vector Machine (SVM), Vikramaditya Jakkula, School of EECS, Washington State University, Pullman 99164.

[11] Ron Meir, Nahum Shimkin, Shie Mannor, ―Machine Learning Lecture Notes #7: Linear Classification Methods, Electrical Engineering Dept., Technion – IIT, 2009.

[12] J. Rekha, J. Bhattacharya and S. Majumder, "Shape, Texture and Local Movement Hand Gesture features for Indian Sign Language Recognition", IEEE International Conference on Trendz in Information Sciences and Computing TISC 2011, 8 - 9th December, 2011, Chennai, India.

[13] William S Noble, "What is a support vector machine?", Nature Biotechnology 24, 1565 - 1567 (2006) doi:10.1038/nbt1206-1565, 12 Dec 2006.