ISSN (e): 2250 – 3005 || Volume, 07 || Issue, 12 || December – 2017 ||
International Journal of Computational Engineering Research (IJCER)

Enhancements in Classfication Analysis Using Support Vector Machine: A Survey

[*1]Dr.J.Chockalingam M.Sc.,M.B.A.,M.Phil.,Ph.D.,
[2]B.Venkatesan M.C.A.,M.Phil.,(Ph.D)
[1]*Associate Professor Research Supervisor, Computer Science Department Khadir Mohideen College, Adirampattinam-India*
[2]*Lecturer Govt.Arts College(A),Kumbakonam Research Scholar, Computer Science Department Khadir Mohideen College, Adirampattinam-India*
*Corresponding Author: *Dr.J.Chockalingam M.Sc.,M.B.A.,M.Phil.,Ph.D.,*

ABSTRACT
Nowadays, Analysis is the most needed machine learning technique for the decision making process in web data and huge amount of data. To perform the analysis in data mining, the some of the techniques such as clustering, classification and generalization are required. The most important process in the data mining is classification because the analysis is required to accomplish the decision making. The classification can be done by the several different techniques such as Case Based Reasoning, Naïve Bayesian Classification, Decision Tree Induction and Support Vector Classification. Among these above, SVM Classification is the most frequently used technique because it performs the analysis in good and expedite as well as in efficient manner. In existing works, many authors proposed and developed their framework with the modified enhancements in the SVM classification based on their application. In this paper, the enhancements in the existing works are discussed and reviewed the best way of enhancements in this algorithm to accomplish the crime factor analysis because it seems to be large. This paper helps to developers to make their decision for using the enhanced SVM in large data applications.
Keyterms: Clustering, Classification, Generalization, Case Based Reasoning, Naïve Bayesian Classification, SVM Classification

I. INTRODUCTION

In today's internet world, the data are collected as in a single data centre for performing analysis through the machine learning classification techniques [1] [2] [4] [5]. Particularly the decision making needed based on the collected data are centralized and analyzed to predict the results. The data mining machine learning techniques were introduced to avoid the time required for the data analysis. Thus the main processes of data mining techniques are clustering, classification and generalization.

Classification is used to analyze the training data based on the test data correlations and identify the patterns to make decisions over the analysis. In earlier machine learning techniques[1] [2] such as Decision Tree, K Nearest Neighbor Classifier is used for obtaining the data classification. However so many existing techniques produces the decision making results for the critical data, the more effectiveness of the data analysis required for the heterogeneous data. Next to the usual machine learning algorithms, the supervised learning models were proposed to fulfill the drawbacks in the machine learning algorithms.

The supervised learning models [1] [2] [5] such as Naïve Bayesian Classification and Support Vector Machine. In this above models, SVM works better than the Naïve Bayesian Classification. Support Vector Classification is used to find the prediction results through the classification and regression analysis. In this paper, the discussions made for the overview of the SVM and its types. There are lot of enhancements can be proposed to accomplish the new grievances in the data regression analysis. Thus the new enhanced SVM can be analyzed to survey the models.

II. SUPPORT VECTOR MACHINE

Support Vector Machine Classification Algorithm [1] [2] [18] is developed by Vapnik and Chervonenkis in 1963. Later 1992, Vapnik, Bernhard and Guyon proposed with the new suggestions to create the nonlinear classifiers by applying the kernel trick to maximum margin hyperplanes. The next evolution in 1995 by Cortes and Vapnik called as current standard incarnation i.e., Soft Margin Linear SVM.

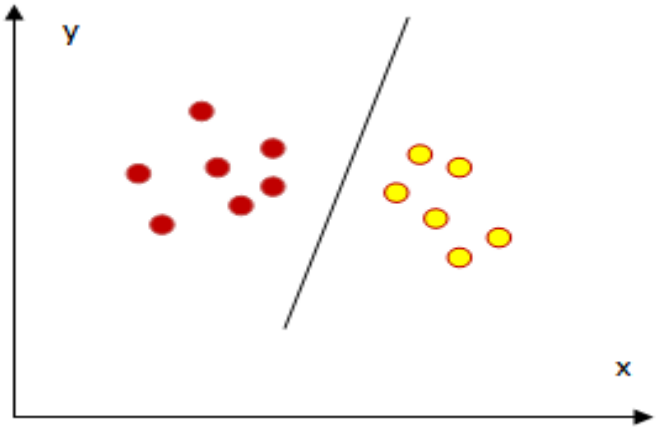

Fig 2.1 Hyper plane with Support Vectors

The SVM based classification [19] is more attractive, because of its efficiency does not depends directly on the dimensions of entities classified. Though SVM is most accurate ad robust but computationally expensive because it is based on convex quadratic programming. SVM can also be reveals the non-linear decision functions by the projection of input test data into a training data in a high dimensional feature space using kernel functions and also formulates the linear classification problem in that training feature space.
The SVM is majorly divided into two types. 1. Linear SVM classifier and 2. Non-Linear SVM classifier. In linear SVM classifier model, the training data samples plotted in the hyperplane space as shown in Fig.2.1. The apparent gap represents the separation of data points. It can be divided into two classes and predicts a straight hyperplane called as a maximum margin hyperplane. In Non-Linear SVM classifier, the training data is generally dispersed up to some extent. Here the kernel trick applies to maximum margin hyperplanes for regression analysis.
The Linear SVM classification [2] can be sub divided into two parts. They are Soft Margin and Hard Margin SVM Classifier. In Soft Margin the feature space in the hyperplane based on the training data cannot be extended. If it is extended type of hyperplane plotting, then the Hard Margin SVM Classifier may used.

A. Limitations of SVM

- The Kernel functions selection is most difficult task in the SVM and it seems to be diplomatic.
- The speed and size is proportional to each other. If the size of the training dataset may increases, then the evaluation speed of the computation may decreases.
- The optimization or enhancement is needed to obtain the SVM Classification in multi dimensional parameters.

III. ENHANCEMENTS IN SVM

It is proven that the SVM plays their role in effective manner but the enhancements are required based on the situation and requirements of the application. The limitations discussed in preceding section are to be overcome through the enhancements achieved in the SVM. The optimization techniques also used for the enhancements. The some of the enhancement techniques discussed below to survey about the enhancements.

A. Multi Class – Support Vector Machine

SVM only originally developed with the binary classification i.e., it have only two classes. It can be extended to multi dimensional because of in need of multiple classes. There are two approaches were proposed for multi class SVM [18] [19]. First is to construct and combine all several binary classifiers and another one is considering all data in single optimization. It is mainly used for the optimization of large amount of data.

B. Hybrid Support Vector Machine

The usual hybrid technique [13] [14] [15] can be takes place in the combination of other techniques with the SVM technique. Most combination is obtained to optimize the SVM technique with the optimization techniques such as Ant Colony Optimization, Genetic Algorithm and Particle Swarm Optimization. Rough set theory also can be used to combine with the SVM to achieve new kind of attribute reduction algorithm by constructing generalized decision information table.

C. Euclidean Support Vector Machine

The SVM classification can be obtained through replacing the optima separating the hyperplane as the classification decision making function based on a distance measurement technique using the Euclidean distance function [10] [11] [12]. As usual the original vector spaces are mapped by means of data point obtained through the average distances of the data point in classification phase.

D. Cluster based Support Vector Machine

The most important enhanced technique [16] proposed to achieve the clustering and classification in a single framework. The data points collected should be clustered through the clustering techniques in data mining are K-Means Clustering and Hierarchical Clustering. After that the clustered vector points are plotted in the hyper plane as multi dimensional class and classify the data through the SVM.

IV.REVIEW ABOUT THE SVM AND ITS ENHANCEMENTS

In existing the authors tried Support Vector Machine with variations in optimizations and enhancements through the combination of different techniques. In this section, these papers are reviewed to conclude the techniques which are best in ever proposed frameworks. In [7], Himani Bavaskar and Mahesh reviews and describes the data classification techniques and widely used algorithms are surveyed. In that paper, they have suggested a robust tool for classification and regression in noisy, complex domains. The two approaches for classification is obtained one is kernel function and another one is linear regression. Fang et.al. [8] proposed sentiment vector space model for coarse grained classification of reviews. It reveals the techniques with the major problems such as selection of classification, feature selection and contribution of features.

The optimized SVM [9] can be proposed with the filter wrapper based gene selection algorithm to classify the microarray data. He wavelet neural networks may kindly used to obtain evaluate the optimized algorithm. This can be generated with multiple parameters. The Euclidean distance [10] [11] calculation to form conventional SVM Classification and achieved high accuracy in the kernel function selection. An enhanced SVM propose with Dynamic Time Warping distance measure has been used as a feature for SVM Classifier [11]. Ping Feng et.al. [12] proposed an enhanced support vector machine model that incorporated the integration of abilities of data preprocessing, parameter selection and rule generation. In this paper, the meta- heuristics rule can be generated to perform the selection of SVM Model. In [13], Rama and Anusha proposed a new multi kernel hybrid support vector machine data classification algorithm to classify the data and compared the other machine learning techniques with the SVM.

A Hybrid Artificial Immune System [14] has designed for health benchmark selection algorithms with Particle Swarm Optimization based SVM with optimizing SVM parameters. Rough set theory is an tool for vagueness and uncertainty information. Here [15] hybrid SVM classification algorithms were proposed to extract the classification rules and also concentrates on generalization of decision information table with new kind of attribute reduction algorithm. Density based SVM [17] also proposed to obtain the results through the benchmark of population density of the data and it also detect the outliers based on linear and non linear SVM techniques.

The cluster based SVM [16] provides the multi dimensional search space data with effective in accuracy and speed. It can be proposed with the K Means Clustering technique and meanwhile improves the ability of generalization of K Means Clustering.

IV. CONCLUSION

The SVM Classification widely used in the decision making applications based on the identified patterns from the training space. In this paper, the Support Vector Machine features and required parameters are analyzed. Thus the survey helps to describe the various enhancements in the Support Vector Machine. It gives the reasons and contributions for the optimization in SVM and other enhancements.

REFERENCES

[1]. Han. J, Kamber.J, and Pei.M, "Data Mining: Concepts and Techniques", Elsevier 2nd Edition.
[2]. Chih-Wei Hsu, Chih-Chug Chang and Chih-Jeh Lin, "A Practical Guide to Support Vector Machine Classification", Dept. of Computer Sci., National Taiwan Univ., Taipei, 106, Taiwan, http://www.cse.ntu.edu.tw/~cjin 2007.
[3]. N.Chirstianini and J.Shawe-Taylor, "An Introduction to Support Vector Machine", Cambridge University, University Press, 2000.
[4]. Bhumika.M and Vimalkumar.B, "Sentiment Analysis using Support Vector Machine based on Feature Selection and Semantic Analysis", Int. Journal of Computer Applications, Vol.146, No.13, July-2016.
[5]. Bholane Savita and Deipali Gore, "Sentiment Analysis on Twitter Data using Support Vector Machine", Int. Journal of Computer Science Trends and Technology, Vol.4, No.3, Jun1 2016.
[6]. Gaurangi P, Varsha G, Vedant K and Kalpana D, "Sentiment Analysis using Support Vector Machine", Int. Journal of Innovative Research in Computer and Communication Engineering, Vol.2, No.1. January 2014.

[7]. Himani Bhavsar and Mahesh H.Panchal, "A Review on Support Vector Machine for Data Classification", Int. Journal of Advanced Research in Computer Engineering and Technology, Vol.1, No.10, December 2012.
[8]. Fang Luo, Cheng Li and Zehui Cao, "Affective feature based sentiment analysis using SVM classifier", IEEE Int. Conf. on Computer Supported Cooperative Work in Design, May 2016.
[9]. Hemalatha.M and Kavitha.S, "Enhanced support vector machine algorithm for classify microarray data", Research Gate Citations, January 2015.
[10]. Lam Hong Lee, Chin Heng Wan, Rajprasad Rajkumar, Dino Lsa, "An enhanced Support Vector Machine Classification framework by using Euclidean distance function for text document categorization", Applied Intelligence-Springer, Vol.37, No.1, pp.80-99, July 2017.
[11]. Thapanan, Phongsakorn S, Haemwaa S and Chotirat A R, "An Enhanced Support Vector Machine for Faster Time Series Classificatiom", Asian Conf. on Intelligent Information and Database Systems, pp.616-625, 2016.
[12]. Ping Feng Pai and Ming Fu Hsu, "An Enhanced Support Vector Machines Model for Classification and Rule Generation", Computationa Optimization, Methods and Algorithms, pp.241-258.
[13]. Anusha R and Dr.N.Rama, "A New Multi Kernel Hybrid Support Vector Machine Data Classification Algorithm", Int. Journal of Scientific and Research Publications, Vol.6, No.12, December 2016.
[14]. Kuo.R.J and Chen.C.M, "Evolutionary based Support Vector Machine", IEEE Int. Conf. on Industrial Engineering and Engineering Management", December 2012.
[15]. Lai Sheng Wang, Yi Tian Xu and Li Shan Zhao, "A kind of hybrid classification algorithm based on rough set and support vector machine", Int. Conf. on Machine Learning and Cybernetics, 2005.
[16]. Yukai Yao, Yang Liu, Yongquig Yu, Weiming Lv, Zhao Li and Xiaoyun Chen, "K-SVM: An Effective SVM Algorithm based on K-Means Clustering", Journal of Computers, Vol.8, No.10, October-2013.
[17]. Zahara Nazari and Dongshik Kang, "Density Based Support Vector Machines for Classification", Int. Journal of Advanced Research in Artificial Intelligence, Vo.4, No.4, 2015.
[18]. Hetl Bhavasar and Amit Ganatra, "Variations of Support Vector Machine Classification Technique:A Survey", Int. journal of Advanced Computer Research, Vol.2, No.4, December 2012.
[19]. Chih-Wei Hsu and Chih Jen Lin, "A Comparision of Methods for Multi-Class Support Vector Machines", Dept. of Computer Sci., National Taiwan Univ., Taipei, 106, Taiwan.

International Journal of Computational Engineering Research (IJCER) is UGC approved Journal with Sl. No. 4627, Journal no. 47631.

*Dr.J.Chockalingam M.Sc.,M.B.A.,M.Phil.,Ph.D. "Enhancements in Classfication Analysis Using Support Vector Machine: A Survey." International Journal of Computational Engineering Research (IJCER), vol. 07, no. 12, 2017, pp. 29-32.