International Journal of Recent Engineering Research and Development (IJRERD)
ISSN: 2455-8761
www.ijrerd.com || Volume 03 – Issue 09 || September 2018 || PP. 01-05

Comparative Study on SVM Multi-class Incremental Learning Algorithms

Divya M R*, Meghana Kishore Murthy*, Shwetha S#

[*] 3rd year student, Information Science and Engineering, R.V.College of Engineering, Bengaluru, India
[#] Assistant Professor, Information Science and Engineering, R.V.College of Engineering, Bengaluru, India

Abstract: Support Vector Machines (SVMs) have been an inherent tool in the field of regression analysis and classification problems, especially in the field of machine learning. SVMs can be extended to various multi-class scenarios in order to enhance the algorithm on which the system works. In this paper we have compared various multi-class algorithms like instance based multi-class learning algorithms, ISOMAP algorithm, class-incremental learning algorithm, context-clustering algorithms, multi-class active learning and cooperative learning algorithms with respect to various parameters like accuracy, compatibility and cost of the training sets. Different applications of these algorithms have been studied, which helps us to understand the importance of SVMs and multi-class learning algorithms in the field of classification.

Keywords: SVM, Multi-class, Classification

I. INTRODUCTION

Support vector machines (SVM) are supervised learning models that are used for regression analysis and classification by analyzing data using different learning algorithms whose general representation is shown in Figure 1. Multiclass algorithms classify objects or instances in to more than one class. The most common form is binary algorithms which classify objects in to two classes. Some algorithms permit the usage of three or more classes. Common approaches to solve multi-classification problems using SVMs are 1-against-rest, 1-against-1, Directed Acyclic Graph SVM (DAG-SVM) and Binary Tree SVM.

The One-against-rest strategy trains exactly one classifier for each class and considers samples of that class as positive and all the others as negative. Despite this approach being popular, a disadvantage that learners see is an unbalanced distribution due to larger set of negatives than the set of positives The 1-against-1 technique trains P(P-1)/2 binary classifiers for a P-way multiclass problem. At the time of prediction, a voting scheme is applied to each classifier after being subjected to a random sample. The class receiving the highest number of positive predictions gets predicted by the combined classifier. This heuristic presents various impreciseinterpretations. The DAG-SVM constructs P(P-1)/2 classifiers for a P-class problem, one for each pair of classes. This approach helps recover weaknesses from previous methods by making the decision faster and more accurate. Binary Tree of support vector machines are used to decrease the number ofbinary classifiers without increasing the complexity of the original program. It is faster than DAG-SVM due toits Log complexity. It is especially useful in problems having a big class number.

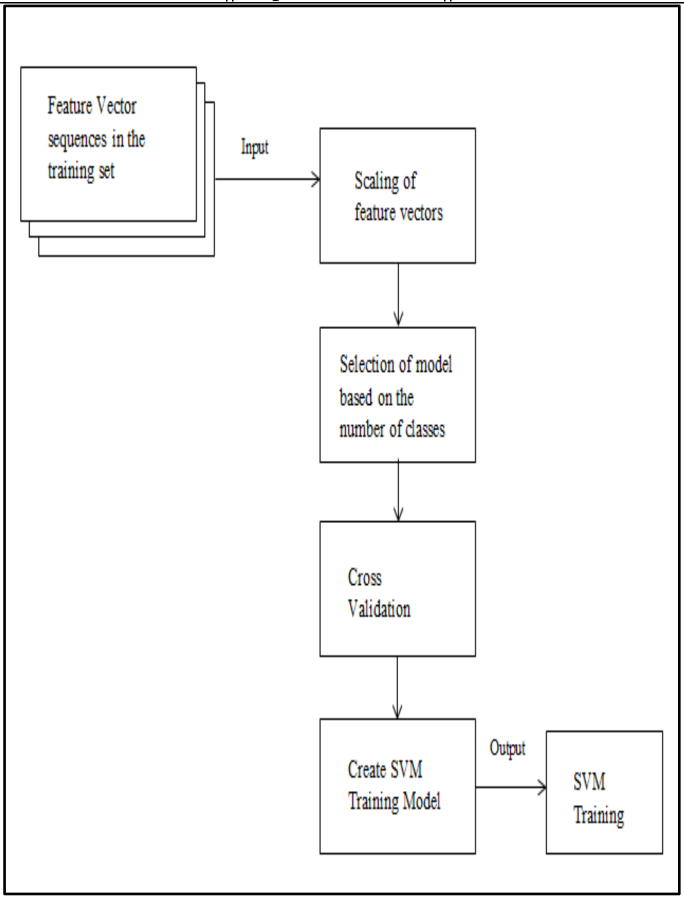

Figure1. General Representation of a SVM model

Every day, the amount of data generated increases vastly and machine learning is becoming largely important. Models are now able to adapt independently on exposure to these new data sets. The rest of the paper discusses various fields in which SVMs are used for multiclass learning algorithms, proving that support vector machines play a vital role in solving an ever-expanding array of classification and regression problems.

Multi-class Active Learning algorithm

Image classification is a very prominent problem in the area of machine learning and pattern classification. This uses various learning techniques which results in exorbitant budget, time and effort. Multi-class active learning [1] is considered as one of the most efficient technique for image classification. It differs from other learning techniques in the way that it relies on uncertainty sampling [2]. It chooses the samples which are not labeled and hard to classify using traditional methods.

Multi-class active learning algorithms helps in decreasing the number of training sets required for the categorization. This learning algorithm selects functional examples for the labeling in the place of meek data. When this algorithm is applied to Support Vector Machines (SVM) which is trained on some training sample, it can be concluded that the classification surface doesn't change even though the entire data set, except the functional vectors, are excluded. It exploits the fact that active selections of labels define the separating surface rather than the data which is redundant to the classifier which holds utmost importance in image classification techniques. Active learning methods for SVM can handle multiclass problems efficiently without the prior awareness of the number of classes, hence promoting incremental approach.

The consequences of training the SVM based on the probability estimation pool using a sigmoid function [3] has been considered. Sigmoid function decides the threshold of the SVM based on the method of class prediction by using the lower limits of the class. The estimation the binary probability of the given dataset and finding the unique global minimum has been analyzed. This can be computed using the method of Gaussian Elimination. The uncertainty sampling uses one more feature for classification known as the entropy measure [4] which is a measure of the imprecise interpretation of the random unknown variable. It has been established that greater the entropy values, greater the uncertainty in the given dataset. Class membership probabilities for the training set are calculated in the active pool and the entropy measure for each of the sample is computed. The measure with the lowest values takes the input and produces the desired result.

The algorithm is robust, helps in easy computation and can be subjected to many number of classes and large data sizes effectively. Although these advantages may make it suitable for image classification, the optimization problem gets convex and complicated due to the limited training set available. By analyzing the algorithm efficiently it can be concluded that the algorithm easily handles the difficulties which arise in multi-class problem statements, works without knowing the cardinality of classes and is interactively efficient with minimal consumption of human time.

An ISOMAP Algorithm based on multi-class multi-manifold learning techniques

The ISOMAP algorithm [5] is one of the most widely used isometric mapping methods which uses low-dimensional embedding methods on a set of high-dimensional data points. It provides simplistic methods to estimate the embedded geometry of a space based on a rough estimate of the neighbors' data points on the data manifold. When there is application of ISOMAP techniques to a multiple class multiple manifold learning algorithm, the result is an intrinsic low-dimensional embedding structure of the given data manifold. This algorithm has been applied to various face recognition systems.

Consider the example of a 64*64 image which can be represented in a 4096 dimensional Euclidean space as a real vector. The results suggest a huge dimensionality space due to which dimensionality reduction techniques like Linear Discriminant Analysis, Generalized Discriminant Analysis must be used.

Though many manifold learning algorithms already exist, such as the classical ISOMAP algorithm, locally linear embedding [6], laplacian eigenmaps [7], they have been implemented only on single manifold data. But in real-time applications, the data is complicated and lies around more than a single manifold. Hence this algorithm effectively exploits multi-manifold data and brings about intrinsic decision making capabilities.

This algorithm is establishes one neighborhood graph over all the points in the multi-manifold data. So, short-circuit edges from all the neighborhood graphs within the data multi-manifold are generated. Later, the data points are selected pairwise between two endpoints of the training set graph. This will result in low-dimensional embedding of data which is the most effective algorithm for systems which involve classification of facial features.

Although the algorithm surpasses many problems of the traditional classification techniques for systems performing facial recognition and classification, it results in difficulties in the computational cost due to the pairwise method mentioned earlier. Also, problems are associated with the algorithm in terms of cost, tuning and many other error-handling parameters.

Hence this algorithm will overcome the disadvantages of the classical ISOMAP algorithm whichonly works on the neighborhood graphs of the dataset. This also helps in obtaining low-dimensional embedding of data whichis a significant improvement over the classical ISOMAP algorithm thus enhancing its features.

Instance Based Multi-class Learning Algorithm

The performance and efficiency of multi-class SVM's can be largely improved by storing and making use of specific instances [8] of the given training set. This algorithm generates the predictions of classification problems based on various instances of the dataset. It uses the concept of decision trees in order to obtain greater accuracies in problems which involve high order classification. The algorithm is based on the concept of using the nearest neighbors in a classification.

In this algorithm, the main process is mapping a given function into various categories based on the attributes associated with the classification.

Consider a feature vector x which belongs to a training set T. Consider a point which belongs to the feature vector x which matches the similarity index with the given training set. It is found that greater the similarity index, greater the accuracy with which the classification of the features are performed. Later, based on the instance decision to which class the feature vector belongs to is made. Also, the noise level of each attribute is calculated. This is storage- specific to the algorithm.

In this algorithm, when a random arbitrary value is chosen based on the similarity index it is accepted only if it has maximum similarity index and it must have a significantly low errorrate. It helps in maintaining the dimensionality of s dataset by making nearly accurate predictions of the feature vectors in the dataset.

This algorithm has many advantages like comprehensibility which indicates that it can be applied to most of the datasets used in real-time applications. Another added merit of this algorithm is that it has a simple implementation phase which leads to complete and consistent verification and validation of the dataset. It is highly robust as it can support various kinds of applications with respect to many fields. They have faster learning rates than most of the traditional algorithms which use the similarity index and the cost of training the dataset is very low since each iteration of the algorithm depends on the current distance of the problem. Though

this approach is used in various approaches of multiclass SVM's, it suffers from abstraction and data sensitivity issues. The requirements related to the storage of the algorithm and rate of learning are inversely related to each other. Due to this rigorous operations on the dataset may lead to frequent system faults.
This algorithm is mainly used in probability models due to its high accuracy and various sensor devices due to the instance specific attribute of the algorithm.

Context Clustering Multi-Class learning algorithm

Despite the Deep Neural Network (DNN) based Statistical Parametric Speech Synthesis (SPSS) system improving modeling precision of statistical attributes, its synthesized speech is not up to the mark as global characteristics of the training data is considered exclusively and not any local characteristics. A context clustering algorithm based on DNN categorizes the training set in to many classes. The Multi-Class Learning(MCL) procedure models the global characteristics as well as those characteristics that are class-dependent of phonetic data thus proving to be advantageous. It prevents the over-fitting difficulties and decreases over-smoothening problem. It has been proved thatthis algorithm has a higher performance than the conventional methods previously used.

DNN bases SPSS systems provide a good standard of synthesized speech when compared to the traditionally used Hidden Markov Models (HMM) [9]. Due to two reasons, the speech may still sound muffled. The first reason is, one network is not enough to seize the vast characteristics of phonetic data, so attributes that have phonetically distinct natures are statistically averaged during the phase of training the model. This results in muffled speech. The second reason is, the structure is limited and does not allow the estimation of the wide variance of parameters.

This approach describes a machine containing two networks. The first network is a context clustering network. The linguistic features are considered as the input and acoustic features are considered as the output, which are clustered into many classes to help prevent over-smoothening the speech attributes. Here, the bottleneck layer is employed as a classifier. The bottleneck layer is trained to represent statistical characteristics of the input and output features and thus is able to effectively estimate the most probable class for any given circumstance. The second network, a MCL network is used to train each input-output pair. The training uses a multi-task learning method i.e. many related tasks are trained at the same time and benefit from each other. Class-independent characteristics are modeled using the shared hidden layers and class-dependent characteristics are modeled using the regression layers.

Cooperative learning algorithm

This cooperative learning algorithm [10] contains two learning algorithms and a sub-optimization problem which is solved by each algorithm (one for bias and one for the support vector). This algorithm is discrete time and hence has a higher convergence speed when compared to those that are continuous time. The one step procedure to classify multi-class data, all the data is trained for learning is compared. Results have confirmed higher performance levels of this introduced algorithm.

The multi-category classification problem is constructing a decision function that distinguishes points in the training data set into different regions so that every region contains points which are a part of all or nearly all of the same set. The cooperative learning algorithm is described briefly as follows: given a kernel function K and training data (which is the input set), calculate the support vectors by using the required iterative algorithm I. Then compute the bias and decision function on the basis of the functions of K classification.

When this algorithm is compared with traditional M-SVM algorithm like M-LSSVR (least square support vector regression) algorithm and Cooperative Neural Network (CRNN) using the three-class Iris data set and three-class Wine data set, it is seen that the algorithm has higher correct rate compared to the others. THE CRNN also obtains a high correct rate, however the speed is lower than half the speed of the improved algorithm. Thus, it can be concluded that the newly introduced algorithm has a high efficiency and good performance in multiclass SVM learning process.

A Class-Incremental Learning algorithm

The Class-Incremental Learning (CIL) method[11] contains two distinct phases: incremental training and incremental feature selection. It has been shown that CIL method has a significantly lower training time than other learning methods, but is also best in terms of performance and effectiveness. Due to the increase in availability of digital text documents, categorization and segregation of data based on machine learning concepts is now a very prominent field of research. Some advantages of this method when compared to popular SVM methods like 1-against-rest, 1-against-1 and divide-by-2 include accuracy are it being faster in training time and significant saving in terms of labor power.

When a new class is added, this method takes all the previously existing classes as one new negative super class against it and uses the basic binary training method on them to get a higher level classifier. When a new class K+1 unknown to the current classification machine is considered, it is not required to train a classifier for all the K+1 classes from the start but only develop a new binary classifier that takes all the samples from known classes as negative and the new unknown class as positive. Using this method, a new class can be incorporated in to the old classification system in an incremental manner.

The main objective of CIL is to reutilize the previously used models to the maximum extent. Incremental feature selection is also done because feature sets are replaced whenever a new class is incorporated in to the classification system. In terms of performance, the CIL method proves to be far better than other commonly used SVM methods and this is largely due to the reutilization of the previously used models. CIL can be used for SVM algorithms but also any other binary classification models by reusing old models.

II. CONCLUSION

Every multi-class algorithm learning algorithm has pros and cons, based on the field in which it used. Different multi-class algorithms have been studied with respect to the application in which they prove to be of most use. In the process of finding the most effective multi-class algorithm for a particular application, the new samples and the previously used support vectors have been examined thoroughly. The multi-class active learning algorithm is used to decrease the number of training sets required for categorization. The ISOMAP algorithm is used to estimate the embedded geometry of a space based on a rough estimate of the neighbors' data points. The instance based algorithm increases efficiency by storing and making use of specific instances of the given training set. The context clustering algorithm categorizes the training set in to many classes which improves performance when compared to conventional methods. The cooperative learning algorithm contains two sub-algorithms which allows higher performance levels. A class-incremental algorithm contains two phases which helps decrease the training time and improve effectiveness. Thus, it can be concluded that the use of multi-class algorithms and support vector machines is an efficient way to solve various problems of classification and regression, each algorithm proving to be useful in its own way.

References

[1]. Ajay.J.Joshi, Fatih Porikili, Nikolaos Papanikolopoulos : Multi-Class Active Learning for Image Classification,*IEEE Explore*. 18 Aug 2019, ISSN 1063-6919

[2]. Yi Zhang, 10-701, Machine Learning, *Spring* 2011 , April 20th, 2011

[3]. Han, Jun; Morag, Claudio The influence of the sigmoid function parameters on the speed of active learning. *In Mira, José; Sandoval, Francisco. From Natural to Artificial Neural Computation.* pp. 195–201 , 2005.

[4]. A. Holub, P. Perona, and M. Burl : *Entropy-based active learning for object recognition*. In CVPR, *Workshop on Online Learning for Classification*, 2008.

[5]. Qicai Cheng, Hongyuan Wang, Yan Feng, Suolan Liu, Lei Xue : A Multi-Class Multi-Manifold Learning Algorithm based on ISOMAP, *IEEE Explore.* 04 Dec 2009, Print ISBN : 978-1-4244-4199-0

[6]. S.T. Roweis, L.K. Saul : Nonlinear dimensionality reduction by locally linear embedding. *Science*, vol. 290, no. 5500, pp. 2323−2326, 2000.

[7]. M. Belkin, P. Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, vol. 15, no. 6, pp. 1373-1396, 2003.

[8]. David W Aha, Dennis Kibler, Mark K Albert : Instance-Based Learning Algorithms, *Springer*, Department of Information and Computer Science, University of California, lrvine, CA 92717

[9]. Eunwoo Song, Hong-Goo Kang : Multi-class Learning Algorithm for Deep Neural Network-Based Statistical Parametric Speech Synthesis. *European Signal Processing Conference.* 2016

[10]. Youshen Xia, Yu Ying : A Cooperative Learning Algorithm for Multiclass Classification. *IEEE, International Conference on Web Intelligence and Intelligent Agent Technology.* 2010

[11]. Bo-Feng Zhang, Jin-Shu Su, Xin Xu : A Class-Incremental Learning Method for Multiclass Support Vector Machines in Text Classification. *Proceedings of the Fifth International Conference on Machine Learning and Cybernetics, Dalian.* 13-16 August 2006