Modified One-Against-All Algorithm Based on Support Vector Machine

Heena Farooq Bhat
Department Of Computer Science,
Kashmir University, India

Mohd Arif Wani
Department Of Computer Science,
Kashmir University, India

Abstract— This paper presents a modification of Support Vector Machine (SVM) One-Against-All Algorithm for multi-class classification task. SVM is originally a model for the binary-class classification. To allow for multi-class classification, different combinations of various binary sub-classifiers are used. . The main objective of this paper is to compare and modify the One-Against-All method of multi-class classification based on support vector machine. This paper discusses One-Against-All SVM algorithm which is able to improve the classification accuracy without imposing high computational cost. Two versions of One-Against-All SVM algorithms are tested on Iris and Lenses data sets. It shows that the proposed algorithm for multi-class problem produce good results.

Keywords— Support Vector Machine, Classification, Multi-class, Support Vectors, One-Against-All Method.

I. INTRODUCTION

SVM's are supervised learning methods used for classification and regression. It involves analyzing a given set of labeled data so as to predict the labels of unlabelled future data. The purpose of SVM is to separate the data points by computing a hyperplane or a decision function. SVM can model complex, real world problems such as text and image classification, hand writing recognition, bioinformatics and biosequence analysis. SVM have been applied in a variety of domains. The criterion used by SVMs is based on margin maximization between the two data classes. The margin is the distance between the hyper planes bounding each class. By maximizing the margin, we search for the classification function that can most safely separate the classes of solvent and insolvent companies. The threshold separating solvent and insolvent companies is the line in the middle between the two margin boundaries, which are represented as $x^Tw+b=1$ and $x^Tw+b=-1$. Then the margin is $2 / ||w||$, where $||w||$ is the norm of the vector w. The separating hyperplane has to be determined in such a way that the margin between positive class and a negative class is maximized to produce good generalization ability.

SVM algorithms are relatively slow. SVM works well for the binary classification problems but there is still no efficient method for K classes. To allow for K Classes or multiclass classification problems, a number of methods have been proposed and is still an ongoing research issue. For large data sets with many support vectors (data points that are closest to the hyperplane), this problem is unresolved. In this paper we will provide a modification of those algorithms by extending them to large datasets. This section provides a brief description of some methods implemented to solve multiclass classification problem with SVM.

II. LITERATURE REVIEW

SVM is originally designed for binary classification and the extension of SVM to the multi-class scenario is still an ongoing research topic [1]. The conventional way is to decompose the *M*-class problem into a series of two-class problems and construct several binary classifiers. Among the binary settings, we can describe several approaches: the comparison of each class against all the others, known as one-vs-all; the comparison of each class against all the other classes individually, known as all-pairs. A good general strategy called pairwise coupling for combining posterior probabilities provided by individual binary classifiers in order to do multi-class classification [2]. SVM's are based on statistical learning theory and have the aim of determining the location of decision boundaries that produce the optimal separation of classes. In the case of a two-class pattern recognition problem in which the classes are linearly separable the SVM selects from among the infinite number of linear decision boundaries the one that minimizes the generalization error. Thus, the selected decision boundary will be one that leaves the greatest margin between the two classes, where margin is defined as the sum of the distances to the hyperplane from the closest points of the two classes [3]. SVMs have been shown to provide better generalization performance than traditional techniques such as neural networks.

[4] Many supervised classification tasks in a wide variety of domains involve multiclass targets. One frequently used and easy method for solving these problems is to train several o_-the-shelf binary support vector machines (SVMs) classifiers and to extend their decision to multiclass targets by using the one-against-one (OAO) or the one-against-all (OAA) approaches. A vast literature exists on the pros and cons of these two approaches, and a comprehensive review can be found for example in [5]. The sequential minimal optimization (SMO) by [6] considers only two variables in each iteration. Then in each iteration of the decomposition method the sub-problem on two variables can be analytically solved. Hence no optimization software is needed. In order to assign a class to a new sample, we need to compute the

output for all binary classifiers and construct an n-dimensional vector to be compared to each class' string [7]. In the OAA approach, the output value of each competing classifier is used in the decision rule rather than the thresholded class prediction as in the OAO approach. The problem with this OAA decision rule is that every classfier participating to the decision is assumed equally reliable, which is rarely the case. This problem has previously been addressed in [8] where a classifier reliability measure is included in the OAA decision process. The most exciting property of SVM is the easy way it can switch from linear to nonlinear margin. It uses Cover's theorem [9], which states that input data, which are linearly non separable into input space *I* (under certain assumptions) can be mapped to another space (called feature space) *F*, in which data can be linearly separable. Inoue and Abe [10] proposed fuzzy support vector machines, in which membership functions are defined using the decision functions.

ALGORITHMS FOR MULTI-CLASS CLASSIFICATION

In this section we will give a brief description of existing algorithms : One-Against-All and will propose an algorithm for One-Against-All which overcomes the limitation of existing one.

A. One Against All SVM:

This is one of the earliest used implementation where the scenario is to decompose K class problems into a series of two class problems. The dataset is classified into K classes. Therefore, it involves K binary SVM classifiers, one for each class. Each classifier is trained to separate one class from the remaining K-1 classes. A hyperplane is defined by the following decision function:

$$F_k(x) = (w_k.x) + b_k$$

The final output is the class that corresponds to the SVM with the largest margin.
This class is determined by the decision rule of the equation as:

$$K^* = \arg \max f_k(x)$$
$$1 \leq i \leq N$$

One of the advantages of this method is that the number of binary SVM classifiers is equal to the number of classes.

1. **Steps of existing Algorithm**:
- It involves N binary SVM classifiers, one for each class.
- Each binary SVM is trained to separate one class from the rest.
- For each class *k* we determine a hyperplane $H_k(w_k, b_k)$ separating it from all other classes, considering this class as positive class *(+1)* and other classes as negative class *(-1)*, which results, for a problem of *K* classes, to *K* binary SVMs.

A hyperplane H_k is defined by the following decision function:

$$\mathbf{F_k(x) = (w_k.x) + b_k}$$

2. **Algorithm for Classifying One-Against-All SVM:**
- Each data point in the test sample is tested for each classifier by finding margin from the linear separating hyperplane.
- The winning class is the one that corresponds to the SVM with highest output value i.e. the largest decision function value.

B. Modified One-Against-All SVM:

Since this method divides each class from the rest of the classes and constructs only k-1 hyperplanes, but this method does not suit all multi-class problems. This is explained in figure1 below. Here class 1 is divided from the rest of the classes by linear hyperplane H1. Similarly, class 4 is divided from other classes by linear separating hyperplane H3. However, Class 2 cannot be separated from the rest of the classes by a single hyperplane. Therefore, one vs all algorithm will not yield good results. Therefore, we propose a new strategy to address this problem. Our strategy requires finding more than one hyperplane that separates a given class from the rest of the classes. Figure 1 shows that two hyperplanes H1 and H2 are required to separate class 2 from the rest of the classes. These two hyperplanes are obtained using the steps explained below

Steps of proposed solution:
1. Consider the given class i to be separated from the rest of the classes.
2. Check if the given class i can be separated from the rest of the classes by a single hyperplane, if yes then goto step 5.
3. Given a class i, find a class from the rest of the classes that can be separated by a hyperplane from class i.
   This is shown in Figure 2a, where the given class 2 is separated from the class 1 using hyperplane H1.
4. Find the hyperplane that separates the rest of the classes except the class of step 3 from the given class i.
   This is shown in Figure 2b, where the given class 2 is separated from the rest of the classes using hyperplane H2.

5.   Save the hyperplane/s that separate the given class i from the rest of the classes. Increment i. Repeat above steps for new i.

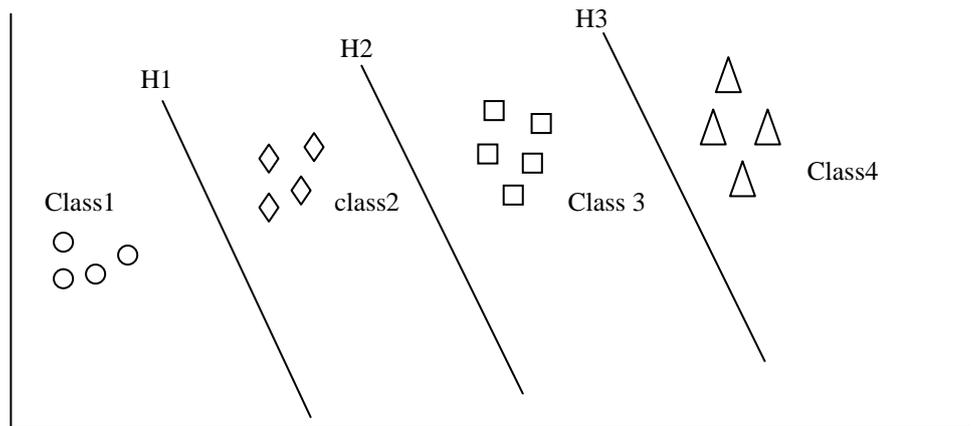Figure 1.

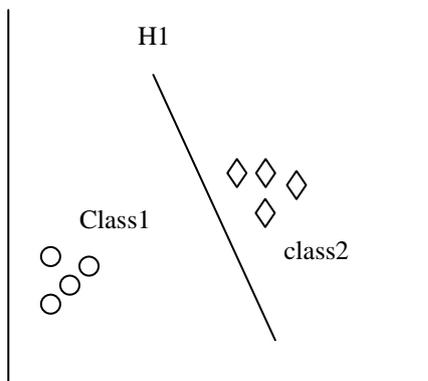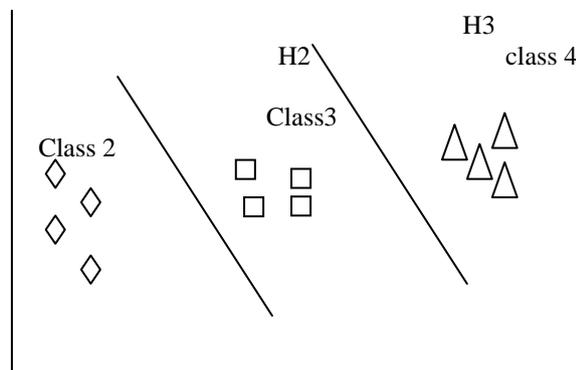Figure 2a.   Figure 2b.

Separation of class 2 from the rest of the classes
Proposed One-Against-All Method

III. RESULTS AND DISCUSSIONS

In this section we will analyze the multiclass classification methods on the basis of accuracy. By comparing the experiment result, the analysis above can be proved.

In this section, we present experimental results on several problems with the following datasets: iris and lenses.
Table 1 and 2 provides the classification accuracy with two multiclass approaches used in present study.

Table 1. Classification Accuracy for Iris dataset

Problem	*Traini ng data*	*Class es*	*Attribut es*	*Accuracy*		
Iris				*Clas s 1*	*Clas s 2*	*Class 3*
OnevsAll	150	3	4	72%	100 %	56%
Modified OnevsAll	150	3	4	80%	100 %	70%
Overall Average acuracy For OnevsAll				76%		
Overall Average acuracy For Modified OnevsAll				83.2%		

Table 2. Classification Accuracy for Lenses dataset

Problem	Training data	Classes	Attributes	Accuracy		
Lenses				Class 1	Class 2	Class 3
OnevsAll	24	3	4	75%	80%	100%
Modified OnevsAll	24	3	4	100%	80%	100%
Overall Average acuracy For OnevsAll				91.3%		
Overall Average acuracy For Modified OnevsAll				95%		

IV.   CONCLUSIONS

Most methods for multi-class classification assume that there is an optimal subset of features that is common to all classes, while in many applications, it may not be the case. In this paper, we compared multiclass support vector machines theoretically and by computer experiments. In this proposed algorithm, some classes use more than one hyperplane in the classification, so it can get higher classification accuracy than the existing method. More elaborated methods allowing the usage of binary classifiers for the resolution of multi-class classification problems are briefly presented. The experimentation of these approaches with SVM's as well as with other learning techniques is a large scale ongoing work and will be presented in the final version of this paper.

REFERENCES
[1]   C.-W. Hsu and C.-J. Lin, "A comparison of methods for multi-class support vector machines", *IEEE transactions on Neural Networks*, vol. 13, pp. 415-425, 2002.
[2]   Hastie, T., Tibshirani, R.: Classification by pairwise coupling. In: Jordan, M.I., Kearns, M.J., Solla, A.S. (eds.): *Advances in Neural Information Processing Systems 10*. MIT Press (1998)
[3]   V.N. Vapnik, "The Nature of Statistical Learning Theory," Springer-Verlag, New York, 1999.
[4]   Kreßel, U.H.G.: Pairwise Classification and Support Vector Machines. In: Advances in Kernel Methods, pp. 255–268. MIT Press, Cambridge (1999).
[5]   Thomas G. Dietterich and Ghulum Bakiri. Solving multiclass learning problems via error-correcting output codes. Journal of Artifcial Intelligence Research, 2:263-286, 1995.
[6]   J. C. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods—Support Vector Learning*, B. Schölkopf, C. J. C. Burges, and A. J. Smola, Eds. Cambridge, MA: MIT Press, 1998.
[7]   Pérez-Cruz, F., & Artés-Rodríguez, A. (2002). Puncturing multi-class support vector machines. In *Artificial Neural Networks—ICANN 2002* (pp. 751-756). Springer Berlin Heidelberg.
[8]   Yi Liu and Y.F. Zheng. One-against-all multi-class svm classification using reliability measures. In Neural Networks, 2005. IJCNN '05. Proceedings. 2005 IEEE International Joint Conference on, volume 2, pages 849-854 2005.
[9]   Cristianini N., J.S. Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods,   CambridgeUniversityPress,2000
[10]   T. Inoue and S. Abe. Fuzzy support vector machines for pattern classification. In Proceedings of International Joint Conference on Neural Networks (IJCNN '01), volume 2, pages 1449–1454, July 2001.