



# Binary tree optimization using genetic algorithm for multiclass support vector machine



Youngjoo Lee<sup>a</sup>, Jeongjin Lee<sup>b,\*</sup>

<sup>a</sup> Manufacturing Technology Center, Samsung Electronics, Samsung-Ro 1-1, Hwasung-Si, Gyeonggi-Do 445-701, Republic of Korea

<sup>b</sup> School of Computer Science & Engineering, Soongsil University, 369 Sangdo-Ro, Dongjak-Gu, Seoul 156-743, Republic of Korea

## ARTICLE INFO

### Article history:

Available online 16 January 2015

### Keywords:

Multiclass support vector machine  
Binary tree architecture  
Genetic algorithm  
Partially mapped crossover

## ABSTRACT

Support vector machine (SVM) with a binary tree architecture is popular since it requires the minimum number of binary SVM to be trained and tested. Many efforts have been made to design the optimal binary tree architecture. However, these methods usually construct a binary tree by a greedy search. They sequentially decompose classes into two groups so that they consider only local optimum at each node. Although genetic algorithm (GA) has been recently introduced in multiclass SVM for the local partitioning of the binary tree structure, the global optimization of a binary tree structure has not been tried yet. In this paper, we propose a global optimization method of a binary tree structure using GA to improve the classification accuracy of multiclass problem for SVM. Unlike previous researches on multiclass SVM using binary tree structures, our approach globally finds the optimal binary tree structure. For the efficient utilization of GA, we propose an enhanced crossover strategy to include the determination method of crossover points and the generation method of offsprings to preserve the maximum information of a parent tree structure. Experimental results showed that the proposed method provided higher accuracy than any other competing methods in 11 out of 18 datasets used as benchmark, within an appropriate time. The performance of our method for small size problems is comparable with other competing methods while more sensible improvements of the classification accuracy are obtained for the medium and large size problems.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

Support vector machine (SVM) is originally designed for a binary classification. SVM classifies a sample as a positive or negative (Cristianini & Shawe-Taylor, 2000; Vapnik, 2000). However, most problems in the real world are usually multiclass problems (Du, Liu, & Xi, 2015; Huang, Zhang, Zeng, & Bushel, 2013; Pathak & Sunkaria, 2014; YJImaz & KJJKcJer, 2013). For multiclass problems, there are more than two classes, and a sample is classified as one class among many classes. SVMs have been shown to perform very well for binary classification problems. In order to deal with multiclass problems, it is necessary to adapt or combine the binary classifiers provided by SVMs.

Several methods have been proposed for solving multiclass problems with SVMs, such as One-Against-One (OAO) (Hsu & Lin, 2002; Kressel, 1999; Platt, Cristianini, & Shawe-Taylor, 2000), One-Against-All (OAA) (Hsu & Lin, 2002; Kressel, 1999; Platt et al., 2000), single-machine (Crammer & Singer, 2002; Hsu & Lin,

2002; Vapnik, 2000; Weston & Watkins, 1999) and direct acyclic graph SVM (DAGSVM) (Platt et al., 2000) methods. The formulation to solve multiclass SVM problems in one steps has variables proportional to the number of classes, thus it is computationally more expensive to solve a multiclass problem than a binary problem with the same number of data (Hsu & Lin, 2002). For a  $k$  class problem, the OAA method constructs  $k$  hyperplanes where each one is constructed by using all data from the training set. The single-machine method also constructs  $k$  hyperplanes similar to the OAA method, however all hyperplanes are obtained by solving one optimization problem. The OAO method constructs  $k(k-1)/2$  hyperplanes where each one is constructed by using the training data from two classes chosen out of  $k$  classes. Although the OAO method constructs  $k(k-1)/2$  hyperplanes, the training time is less than the other methods since each SVM uses a small number of training data for learning. The DAGSVM method also utilizes  $k(k-1)/2$  hyperplanes. However, the testing time is less than that of the OAO method. Hsu and Lin (2002) compared the above-mentioned methods in terms of their performance and computational cost. They concluded that no method can compete with the OAO method in the training time and no method is statistically better

\* Corresponding author. Tel.: +82 2 820 0911.

E-mail addresses: [spicio2@gmail.com](mailto:spicio2@gmail.com) (Y. Lee), [leejeongjin@ssu.ac.kr](mailto:leejeongjin@ssu.ac.kr) (J. Lee).

than the others in the generalization performance. Thus the OAO and the DAGSVM methods are more suitable for practical use than the other methods. However, for all these methods, too many binary tests are required to make a decision. Although DAGSVM may reduce the binary test times, DAGSVM requires  $n - 1$  tests at least when  $n$  is a class number. This shortcoming results in inefficiency of multiclass SVM in applications with many classes.

Recently, the hierarchical structure method has been researched. This method is similar to the output coding method, but it employs binary SVMs step by step according to a hierarchical structure. Tree architecture is always leveraged in decision theory. Support vector machines with binary tree architecture (Cheong, Oh, & Lee, 2004) has been introduced to reduce the number of binary classifiers and to achieve a fast decision. This method took advantage of both the efficient computation of the tree architecture and the high classification accuracy of SVMs. This method used the kernel-based self-organizing map to convert the multiclass problems into proper binary trees by maximizing the scattering measure at the kernel space. However, it required exhaustive search for all possible combinations. Although this conversion is done only once at the classifier design stage, that is poorly acceptable for a real application and an efficient grouping algorithm should be vitally required. In OneAtOnce method (Young, Yen, Pao, & Nagurka, 2006), a one-class-at-a-time removal sequence planning method was proposed to decompose a multiclass classification problem into a series of two-class problems. This method used the fixed binary tree structure which was generated to maximize the depth of a binary tree. Then, the optimization was performed in the fixed tree structure. However, its complexity is proportional to the factorial of the number of classes.

To solve this binary tree design problem, a binary tree design for multiclass SVM using genetic algorithm (GA) is proposed in this paper. Instead of utilizing the clustering method (Chen, Wang, & Shen, 2011) to group different classes together to train a global classifier or determining the classification order in a fixed structure of binary tree, we propose the binary tree optimization method to find a quasi-optimal binary tree design by genetic evolving process. Unlike previous researches (Qin, Qin, Wang, & Lun, 2013; YJlmaz & KJljkçer, 2013) on multiclass SVM using binary tree structures, our approach globally finds the optimal binary tree structure for the improvement of the classification accuracy without the sacrifice of the computational efficiency. A tree structure encoding scheme is developed to represent a tree design in a chromosome, and an enhanced crossover strategy is proposed with the preservation of a sub-tree structure. Our enhanced crossover strategy includes the determination method of crossover points and the generation method of offsprings to preserve the maximum information of a parent tree structure.

The remainder of this paper is organized as follows. The next section explains related works for solving multiclass problems with SVMs. Section 3 describes the proposed method of the binary tree design for multiclass SVM using GA. Section 4 presents the results of the application of the proposed method to benchmark datasets along with a comparison with the results obtained with other methods. Finally, we summarize the results and discuss future work in Section 5.

## 2. Approaches to multiclass problems with SVMs

There are mainly two kinds of approaches for multiclass SVM. The first, output coding method constructs several binary SVMs and simultaneously employs all of them to predict the class of new samples. The second, hierarchical structure method employs binary SVMs step by step according to the hierarchical structure. Following sections briefly introduce these representative methods.

### 2.1. Output coding method

Two output coding methods, OAA and OAO, are currently the most popular strategies to deal with multiclass classification problems. The former decomposes the  $c$  class classification problem into  $k$  binary classification sub-problems. Each classifier separates one class from the remaining  $k - 1$  classes. The class label of the test sample is decided by combining the  $k$  classifier outputs using winner-take-all methods (Anand, Mehrotra, Mohan, & Ranka, 1995). The latter forms a binary classifier for each class-pair and thus  $k(k - 1)/2$  classifiers are required. For the test sample, the decision is made by combining these  $k(k - 1)/2$  classifier outputs using a majority voting strategy (Friedman, 1996). However, in OAA method, the similarity between the classes is not considered so that there is no guarantee that good discrimination exists between one class and the remaining classes. In OAO method, a  $k$  class classification problem is exhaustively decomposed into a set of  $k(k - 1)/2$  classifiers, and the number of classifiers and computations are prohibitive (Kumar, Ghosh, & Crawford, 2002).

The third popular approach is DAGSVM (Platt et al., 2000). Its training phase is the same as OAO method by solving  $k(k - 1)/2$  binary SVMs. However, in the testing phase, it uses a rooted binary directed acyclic graph which has  $k(k - 1)/2$  internal nodes and leaves. Each node is a binary SVM of  $i$ th and  $j$ th classes. Given a test sample, starting at the root node, the binary decision function is evaluated. Then, it moves to either left or right depending on the output value. Therefore, we go through a path before reaching a leaf node, which indicates the predicted class. The advantage of DAG is that (Platt et al., 2000) some analysis of generalization can be established. There are still no similar theoretical results for OAA and OAO methods, yet. In addition, its testing time is less than the OAO method.

### 2.2. Hierarchical structure method

Over the last decade, there has been great interest in reducing the number of classifier to be trained and tested. To deal with it, a binary tree based SVM has been researched. The essential part of binary tree approach to get high accuracy is the design of the binary tree. Hence, many efforts have been made to develop an optimal binary tree design. Binary tree approaches can be briefly categorized into two groups according to the binary tree design method; clustering based method and DAGSVM variations.

Standard  $k$ -means clustering has been applied to design a binary hierarchical structure (Vural & Dy, 2004). To divide a set of classes into two groups ( $k = 2$ ) at each node, they used the class mean  $\mu_i$  to represent each class  $i$  and applied standard  $k$ -means clustering to divide these mean vectors into two groups, such that the total distance between the means  $\mu_i$  (of all classes in cluster  $C_j$ ) to its cluster centroid  $C_j$  is minimized. A similar methodology has been used in Wang and Casasent (2008) where one representative vector of each class (a combination of its support vectors) was used in place of  $\mu_i$ . Instead of  $k$ -means clustering, other clustering methods were also used. In Cheong et al. (2004), a kernel-based self-organizing map (SOM) was used in the conversion of the multiclass problem into binary hierarchies. Benabdeslem and Bennani (2006) introduced dendrogram-based SVM, which utilizes agglomerative hierarchical clustering. This method is a kind of hierarchical clustering method to build the class binary tree.

The class binary tree can be obtained by modifying the graph of DAGSVM. There are two variations on DAGSVM. First one is a decision based OAA (Debnath, Takahide, & Takahashi, 2004). The basic concept is illustrated in Fig. 1. In this figure, the hyperplane  $f_{14}$  is not needed for producing the boundary of class 1, and the hyperplane  $f_{41}$  does not contribute to the boundary of class 4, which is distant from class 1. From this illustration, we can see that

Download English Version:

<https://daneshyari.com/en/article/10322786>

Download Persian Version:

<https://daneshyari.com/article/10322786>

[Daneshyari.com](https://daneshyari.com)