International Journal of Computer Applications (0975 – 8887)
Volume 108 – No. 11, December 2014

Advance Probabilistic Binary Decision Tree using SVM

Anita Meshram
School of Information Technology
UTD, RGPV, Bhopal, M.P., India

Roopam Gupta
School of Information Technology
UIT, RGPV, Bhopal, M.P., India

Sanjeev Sharma
School of Information Technology
UTD, RGPV, Bhopal, M.P., India

ABSTRACT
The probabilistic decision tree to an actual diagnosis database is in progress, where the performance of the probabilistic decision tree is tested in view of the size of the databases and the difficulties is that it implies for processing them. Here proposed an algorithm Advance Probabilistic Binary Decision Tree (APBDT) using SVM for solving large class problem and it performs better when increase the size of the database. APBDT-SVM combines Binary Decision Tree (BDT) and Probabilistic SVM is an effective way for solving multiclass problem. Probabilistic SVM uses standard SVM's output and sigmoid function to map the SVM output into probabilities. Using APBDT-SVM classification accuracy can be improved and training-testing time can be reduced.

Keywords
SVM, Probabilistic SVM, Binary decision tree, separability measures

1. INTRODUCTION
Support vector machine(SVM) has become one of the most widely used machine learning algorithm, mostly for classification[1]. SVM is classifiers which is originally designed for solving binary classification problem and the extension of SVM to the multi-class problem is still an ongoing research issue[2]. The standard SVM multiclass approaches such as "one-against-one"(OaO), "one-against-all"(OaA)[3] or Directed Acyclic Graph(DAG)[4] have shown adequate result when separating the classes, but don't take into account the structure and the distribution of data. To overcome this drawback a simple and intuitive approach came which is based on building a binary decision tree[5].

The structure of decision Tree is constructed by measuring the distance between the gravity centers of different classes, An automated graph is generated where at each node a binary-class SVM is trained.

In a binary-class context platt[6] proposed a method for extracting probabilities *p(class|input)* from SVM output, which is used for classification tasks. The approach follows as, training for the provided parameters make analyse for results. The underlying idea of probabilistic SVM classifier(PSVM) is that the distance from an example of the first one is having larger length, the example is matched to that particular class, which makes that the example similar to existing classes.

1.1. Probabilistic Support Vector Machines
SVM only gives a class prediction output that will be either +1 or -1. There are several approaches have been proposed in order to extract associated probabilities from SVM output. Here will be focused on platt's approach. He has composed a sigmoid function between the output f(x) of SVM and the probability of membership p(y=i|x) to a class i, specified by the attribute x.

Expression for the sigmoid function are:

$$p\big(y = 1\big|f(x)\big) = \frac{1}{1 + e^{af(x)+b}}$$

Where a and b are parameters computed from the minimization of the negative log-likelihood function[6]

$$\min \quad -\sum_i t_i \log p\big(f(x)\big) + (1 - t_i)\log(1 - p(f(x_i)))$$

And t_i is the new label of the classes. the probability of correct label can be derived by using the following rule. The estimate for target probability of positive and negative examples is,

$$t_+ = \frac{N_+ + 1}{N_+ + 2} \qquad\qquad t_- = \frac{1}{N_- + 2}$$

Where N_+ and N_- are the number of points that belongs to class *1* and class *2* respectively.

1.2. From Binary Class to Multi-Class Problem
Support Vector Machine is an original design for binary classification task, the extension for binary classification to multiclass classification is an ongoing research issue. There are different methods to solve the multiclass problem: One-against-One(OaO), One-against-All(OaA), Directed Acyclic Graph(DAG), Binary Decision Tree(BDT).

1.2.1 One Against All(OaA)
OaA[3] construct N two-class SVMs. The i^{th} SVM are trained while i^{th} class is labeled by 1 and rest sample are labeled by -1. In the testing phase, a test example is presented to all N SVMs and is labelled according to the maximum output among the N classifier. The disadvantage of this method is that its training and testing phase are usually very slow.

1.2.2 One Against One(OaO)
It constructs all possible N(N-1)/2 two class classifiers. Each classifier is trained by using the sample of first class is labeled 1 and sample of another class is labeled -1. A max-win algorithm is used to combine these classifiers. Each classifier casts one vote for its recommended class, and finally the class with the highest votes wins. This methods disadvantage is that, when the number of classes is large then OaO resulted slower testing because every test sample has been presented to the large number of classifiers N(N-1)/2.

1.2.3 Directed Acyclic Graph SVM(DAGSVM)
The DAGSVM algorithm for training an N(N-1)/2 classifiers is same as OaO. In the testing phase, the algorithm depends on rooted binary directed acyclic graph to make a decision. So the classification of DAG is usually faster than OaO.

1.2.4 Binary Decision Tree(BDT)

This method uses multiple SVMs arranged in a binary tree structure. SVM in each node of the tree is trained using two of the classes. In this architecture, N-1 SVM needed to be trained for N Class problem, but it only requires to test $\log_2 N$ SVMs to classify a sample. This lead to an impressive improvement in recognition speed when addressing problems with big number of classes.

To build a binary tree here first start by dividing the classes into two disjoint group g1 and g2 as shown in figure 1. This is performed by calculating the N gravity center for N different classes and measuring the distance between the gravity center of different classes. The two classes that have biggest Euclidean distance are assigned to each of the two groups. After that the classes with the smallest distance from one of the group assigned to the corresponding group. The classes from the first group are assigned to left sub-tree and the classes to the second group are assigned to the right sub tree. The process is continuing by dividing each of the groups to its subgroup, until they reached only one class per group which defines a leaf in the decision tree.

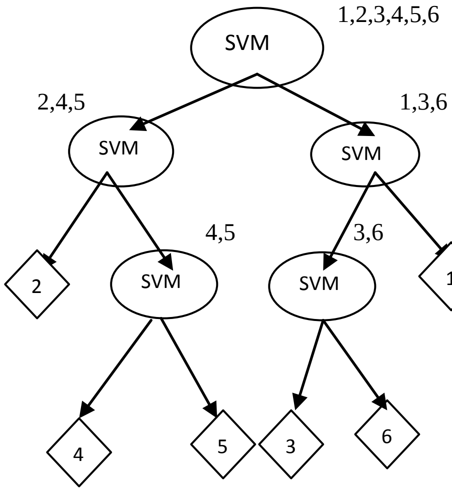

Figure 1: BDT for classification of 6 classes from 1 to 6

2. LITERATURE REVIEW

Madzarov et al.[5] is presented a novel architecture of Support Vector Machine classifiers utilizing binary decision tree(SVM-BDT) for solving multiclass problem. Design the hierarchy of binary decision subtask using SVMs with clustering algorithm. The clustering model utilizes distance measures at the kernel space instead of input space. The experimental result indicates that the training phase of SVM-BDT is faster while comparing better accuracy with other SVM based approaches, ensembles of tree(Bagging and random forest) and neural network. During the testing phase, due to its logarithmic complexity, SVM-BDT is much faster than widely used multi-class SVM methods like OaO and OaA.

Platt et al. [6] present a method for extracting probabilities *p(class|input)* from SVM outputs, which is used for classification post-processing. Standard SVM does not provide such probabilities. After train an SVM, train the parameters of an additional sigmoid function that used to map the SVM output into probabilities. In the Experimental steps SVM+sigmoid yield probabilities are compared to the raw SVM.

Binary tree support vector machine, is an effective way for solving multiclass problem, which combines support vector machine and binary trees. Sun et al.[7] is proposed that, To maintain high generalization abilities, most separable classes should be separated at the upper nodes of the binary tree. A new binary tree with fewest levels is established based on clustering method. Experiment results show loherer decision time and better accuracy with 93.59% by comparing with oblique binary tree, balance binary tree and unbalance binary tree.

Mulay et al. [8] propose decision tree based support vector machine which combines support vector machine and decision tree. It is an effective way for solving multi-class problems. This method decreases the training and testing time, increasing efficiency of the system. They include different ways to construct the binary trees by dividing the data set into two subsets from root to the leaf until every subset consist only one class. Euclidean distance is used for measuring the separability between the classes.

Bala et al.[9] proposed decision tree SVMs architecture which is constructed to solve multi-class problem. In this paper maintain high generalization abilities, by determining optimal structure of decision tree using statistical measures for obtaining class separability. ODT-SVM takes advantage of both the higher classification accuracy of SVM and efficient computation of decision tree architecture. Performance is evaluated in terms of computation time and classification accuracy. The highest accuracy is achieved by 100%. Experimental performance shows that ODT-SVM is significantly better in comparison to conventional OaO and OaA in terms of both training and testing time.

Arun et al.[10] present an improved version of One-against-All(OAA) method for multiclass SVM classification based on decision tree. The proposed decision tree based OAA(DTOAA) aimed, to increasing the classification speed of OAA by use of posterior probability estimates of binary SVM output. DT-OAA decreases the average number of binary SVM test that required in testing phase. It is compared to OAA and other multi class classification methods. Computational comparison indicates that the proposed method can achieve almost the same accuracy as OAA with 99.92% but much faster in decision making.

When there are hundred or even thousands of classes, existing technique for mapping multiclass problem into a simple binary classification problem give serious efficiency problem. Rocha et al.[11] introduce the concept of correlation and joint probability of base binary learners. They also discuss two additional strategies: one to reduce the number of required base learner in multiclass classification, and another is to find new base learner that might be best complement to the existing set. Highest classification accuracy is achieved by 96%. The goal is to keep up the efficiency by finding most discriminative binary classifiers to solve the multiclass problem.

Sidaoui et al.[12] examine the performance of framework for solving multiclass problem with support vector machine(SVM). The proposed paradigm builds a binary tree for multi-class SVM, using criteria of natural classification : Homogeneity and Separation, with the aim of obtaining optimal tree. This proposal is more accurate in construction of the tree. In the test phase, due to log complexity it is much faster than other methods, that have the problem of big class number. Here recognition rate was achieved by 57% on voherels of TIMIT corpus and 97.73% on MNIST dataset for 10 digits. Here training time and number of support vectors are also reduced compared to other methods.

The tree architecture has been employed to solve multi-class problem based on SVM. Most of tree based SVM classifiers try to split the multiclass space by using a clustering algorithm. Cohen et al.[13] presented a preliminary and promising result of a multiclass space partition method, which account for the database class structure and allow the node's parameter specific solution. Each node in the space is split into two class problem possibilities and the best SVM solution are found. Preliminary result shows that accuracy is improved, less information is require and hard separable classes can easily be identified.

Madzarov et al[14]. Propose that hierarchy of binary decision subtask using SVM is designed with clustering algorithm. They investigate how different distance measures for the clustering influence the predictive performance of SVM-BDT. Distance measures that they consider include Euclidean distance, Mahalanobis distance and standardized Euclidean distance. The performance of this architecture is compared with four other SVM based approaches, ensembles of decision tree and neural network. Experimented result suggest that the performance of architecture significantly varies depending upon applying the distance measure in the clustering process.

Table 1: Comparative Analysis

Title	Methods	advantage	Disadvantage
Optimal decision tree based multi-class support vector machine	OaO and OaA Optimal decision tree	- Overcome the problem of unclassifiable region. - Improve generalization ability. - Less computation time.	- Among four measures occupied for determining the structure of decision tree, neither of them declare a clear winner over another in terms of computation time for training and testing.
Fast multiclass SVM classification using decision tree based One against All method	Decision tree based OaA	- Increase classification speed. - Faster in decision making. - Improve generalization with cost. - Give better accuracy result.	- Almost the same classification accuracy as OaA method. - Increasing threshold δ increases the accuracy of DT-OAA but at the same time decreases classification speed.
Evaluation of distance measures for multi-class classification in binary SVM decision tree.	SVM based binary decision tree	- Efficient computation - High classification accuracy - Dramatic improvement in recognition speed when addressing problem with big number of classes.	- When using Mahalanobis distance for big number of classes, it gives poor result in training, testing and prediction error rate.
Multiclass from binary: Expanding one-vs.-all, one-vs.-one, and ECOC based approaches.	One- vs.-All, One-vs.- One, ECOC	- Its implementation is simple. - Highly computational intensive. - Better classification result. - It is fast and does not impact the multiclass procedure.	- It does not hold all cases all time. - Existing methods for binarization of multiclass problem are either impractical or do not fare so well.
A new multi-classification method based on binary tree support vector machine.	Binary tree, types of binary tree.	- Lower number of binary classifiers. - Faster decision speed - No nonseparable region exists.	- Occurrence of classification error in the root node, resulted increases the overall classification error.

Uribe et al. [15] describes an original classification technique, the Probabilistic Decision Tree(PDT) producing a posterior probability in a multiclass context. They describe a method based on a Binary Decision Tree(BDT) with Probabilistic Support Vector Machine classifier(PSVM). At each node of the tree, bi-class SVM along with sigmoid function are trained to give a probabilistic classification output. At the experimental step PDT are tested on benchmark datasets and the result is compared with another multiclass methods. The highest accuracy is achieved with 92.75% with compared to other multiclass methods as OvO, DAG, RL-BDT.

3. PROBLEM FINDING

When there are hundreds or even thousands of classes, existing technique for mapping multiclass class problem into a simple binary classification problem gives serious efficiency problem. Uribe [15] propose Probabilistic Decision Tree using SVM which is original and new classification technique. It gives better results than other multiclass methods like OaO, DAG, and BDT. Here experiment will be performed by taking a small number of classes. So literature [15] not specifies that PDT-SVM also gives better result when using this algorithm in the large class problem. The application of PDT to the actual diagnosis database is in progress, where the performance of the PDT is tested in view of the size of the database and the difficulties that it implies for processing them.

4. PROPOSED SOLUTION

Here proposed that Advance Probabilistic Binary Decision Tree(APBDT) using SVM for solving large class problem. Advance PBDT is an extension of PDT[15]. Proposing advance probabilistic binary decision tree can be used in large class problem and also it better perform when increases size of the database. On the basis of Binary Decision Tree and Probabilistic output of Support Vector Machine here want to present Advance Probabilistic Binary Decision Tree (APBDT) using Support Vector Machine(SVM) as an original approach to the multi-class classification problem. Instead of using a simple SVM classifier in each node, here propose SVM classifier associated with a sigmoid function(PSVM) to estimate the probability of membership to each sub-group in the node. Proposing classifier architecture APBDT-SVM takes the advantage of both the highest classification accuracy of SVM and the efficient computation of the tree architecture.

APBDT-SVM is based on recursively dividing the classes into two groups in every node of the Binary Decision Tree and training an SVM associate with sigmoid function i.e. Probabilistic SVM would decide incoming unknown sample should be assigned in which of the group.

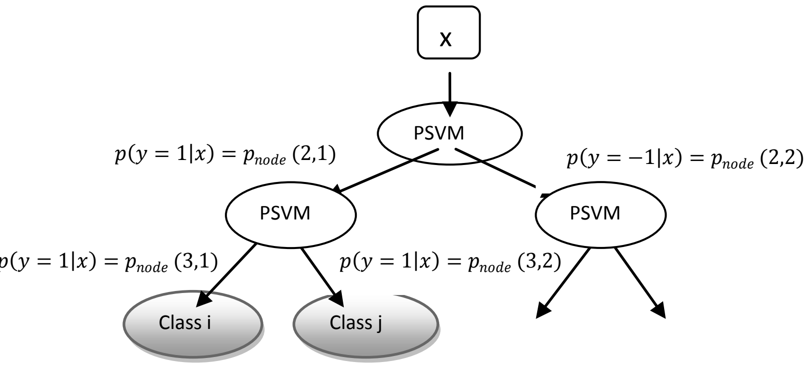

Figure 2: Example of probabilistic binary decision tree using SVM (PBDT-SVM)

One way to get to a leaf node, here unique probability function is used for a trained tree.

$$p(y = i|x) = \prod_{h=1}^{leaf} p_{node}\ (h, l)$$

h is the level of the tree and h=1 is the root node. This expressly state that the probability of membership of an element to the class i is calculated as the product of the probabilities of the decisions taken in all the nodes visited until arriving to the leaf. $node(h, l)$ Here means that *l* node in the *h* level. Once the tree builds here will have the probability function, one for each class. When classifying unknown cases here will just evaluate the probability function and then choose the class with the highest score.

4.1 Algorithm of Advance Probabilistic Binary Decision Tree

Here can understand the complete proposed work through these steps:

Step 1: Training phase

Inputs: Training set

Outputs: Probabilistic functions (one for each class)

1. Build a binary decision tree.
   a. Calculate N gravity center for N different classes.
   b. Calculate Euclidean distance between two class's gravity centers.
   c. Let classes C_i and C_j have biggest Euclidean distance. So the classes which have biggest Euclidean distance assign in two different groups *g1* and *g2*.
   d. Classes which have smallest distance with C_i compare to C_j assign in *g1*, otherwise assign in *g2*.
   e. Go to next step if there is one class per group. Otherwise repeat *step c* and *d*
2. Training an SVM classifier for each node of the decision tree.
   a. Calculate the hyper plane that separate the classes.
   b. If Separated classes are plural class go to *step a*
   c. Else go to the next step.
3. Fit sigmoid function to every SVM classifier trained in *Step 2*. Obtaining a probability function for the node.
4. For each leaf, probability function should transverse all the corners.

End step 1

Step 2: Testing phase

Input: An unclassified example *x*

Output: Records of proposed classes and their probability.

1. Evaluate all the probability functions for the new unclassified example using *step 1(3-4).*
2. Arrange the classes according to the probabilities.

End step 2

By using APBDT-SVM classification ability can be improved, increases the efficiency of the system and training-testing time can be reduced. Error rate of SVM can be also decreased by using binary decision tree.

5. CONCLUSION

The APBDT-SVM is providing superior multiclass classification performance. Utilizing a decision tree architecture with a probabilistic output of SVM takes much less computation for deciding that in which class unknown sample is put down. Here proposed a new and original technique that combines Binary Decision Tree and SVM associate with a sigmoid function(PSVM) to estimate the probability of membership to each sub-group. Probabilistic function for each leaf are built after traversing complete nodes and leaves. It is critical to have proper structuring for the good performance of APBDT-SVM. By using APBDT-SVM classification accuracy can be improved, training and testing time can be reduced, on utilizing APBDT proposed algorithm work can be further used in various classifications of the data on various aspects efficiently when we compare the same with existing mentioned algorithm.

6. REFERENCES

[1] V.N. vapnik; the nature of statistical learning theory; Springer, Newyork, 1995.

[2] C. W. Hsu., C.J. Lin; "A comparison methods for multiclass support vector machines"; IEEE transaction on neural networks, vol. 13, no. 2, pg. 415-425, March 2002.

[3] Y. Liu and Y.F.Zheng; "One-against-all multi-class SVM classification using reliability measures"; IEEE international joint conference on neural network (IJCNN); vol. 2, pg: 849-854, 31 July- 4 Aug. 2005.

[4] J. C. Platt, N. Cristianini and J. Shahere-Taylo;, "Large margin DAGs for multiclass classification"; Advances in

neural information processing system, vol. 12, no. 3, pg. 547-553, 2000.

[5] G. madzarov, D. gjorgjevikj and I. chorbev; "A multi-class SVM classifier utilizing binary decision tree"; An international journal of computing and informatics, Informatica; vol. 33 number 2; ISSN0350-5596, Slovenia; pg:233-241, 2009.

[6] J. Platt; "Probabilistic outputs for support vector machines and comparison to regularized likelihood methods"; in Advances in large margin classifiers, Cambridge, MIT press, 2000.

[7] G. sun, Z. Wang, M. Wang; "A new multi-classification method based on binary tree support vector machine"; 3rd international conference on innovative computing information and control (ICICIC) IEEE, Dalian, Liaoning; pg. 77; 18-20 June 2008.

[8] S. A. Mulay, P. R. Devale,G. V. Garje; "Intrusion detection system using support vector machine and decision tree"; International journal of computer application; vol. 3, Number 3, pg. 0975-8887; June 2010.

[9] M. Bala, R. K. Agrawal; "Optimal decision tree based multi-class support vector machine"; Informatica; school of computer science & system sciences, vol. 35, pg. 197-209; 2011.

[10] M. Arun Kumar, M. Gupta; "Fast multiclass SVM classification using decision tree based one-against-all method"; Springer, neural process lett, vol. 32, pg. 311-323; 25 Nov. 2010.

[11] A. Rocha, S. Goldenstein; "multiclass from binary: Expanding one-vs.-all, one vs. one and ECOC-based approaches"; IEEE transaction on neural networks and learning system; vol. X, number Y; Aug. 2013.

[12] B. Sidaoui, K. Sadouni; "Efficient approach one- versus-all binary tree for multiclass SVM"; Springer link; Transactions on engineering technologies; vol. 275, pg. 203-214; 2014.

[13] D. A. Cohen, E. A. Fernandez; SVMTOCP: "A Binary tree based SVM approach through optimal multiclass binarization"; Springer link; progress in pattern recognition, image analysis, computer vision and application; vol. 7441, pg. 472-478; 2012.

[14] G. madzarov, D. gjorgjevikj; "Evaluation of distance measures for multi-class classification in binary SVM decision tree"; Springer link; Artificial intelligence and soft computing lectures notes in computer science; vol. 6113, pg. 437-444; 2010.

[15] J. S. Uribe, N. Mechbal, M. Rebillat, K. Bouamama, M. Pengov; "Probabilistic Decision Tree using SVM for multiclass classification"; Conference on control and fault-tolerant system. IEEE; pg. 619 to 624; Oct 9-11, France 2013.