A new Ensemble Learning Method Based on Pairwise Constraints and Subset Selection

RanRan Ma\(^1,2\), XiaoShi Zheng\(^1\), YanLing Zhao\(^1\), and ChengZhong Yang\(^3\)
\(^1\) Shandong Computer Science Center, Shandong Jinan, China
Email: maran.ran@163.com
\(^2\) Shandong University of Science and Technology, Shandong Qingdao, China
Email: zhaoyl@keylab.net  fgkm@163.com

I. INTRODUCTION

Ensemble learning techniques is now an active area of research in machine learning because of its better generalization ability than its base classifiers. It trains multiple component classifiers and combines their predictions. Generally, the ensemble structure usually proceeds in two steps: classifiers construction and classifiers combination. In this paper, we focus on the first problem and adopt a simple majority voting scheme to combine the predictions of multiple base classifiers.

It is known that in order to get a strong ensemble, the component classifiers should be with high accuracy as well as high diversity. However achieving such goal is not easy. Kuncheva[1]summarized four fundamental approaches for creating diverse classifier for ensembles: 1) using different data subsets; 2) using different feature subsets; 3) using different algorithm to train base classifiers; 4) using different combining schemes. The first two approaches attract many researchers. They tend to construct base classifiers by sampling the instances or their input features. Bagging[2], Boosting[3] and Random Subspace[4] are three main representatives for these two categories. Bagging generates diverse classifiers by randomly selecting subsets of samples to train classifiers. Intuitively, based on different sample subsets, classifiers would exhibit different behaviors. Boosting used parts of samples to train classifiers as well, but not randomly; difficult samples have a greater probability of being selected, and easier samples have less chance of being used for training. With this mechanism, most created classifiers will focus on hard samples and can be more effective. The Random Subspace method creates various classifiers by using different subsets of features to train them. Because problems are represented in different subspaces, different classifiers develop different borders for the classification.

II. RELATED WORKS

In fact, most ensemble methods proposed subsequently are based on these above three algorithms or jointly combine them. For example, Breiman[5] integrated both bagging and random subspace, using decision tree as the base classifiers. Rodriguez et al[6] proposed the rotation forest method which first randomly partitions features sets into K subsets, performs principal component analysis on each subsets of a bootstrap replicate, and reassembles those PCA projective vectors into a rotation matrix to form new features for a base classifiers. Nicol`as Garc`ia-Pedrajas[7] proposed a nonlinear boosting projection for ensemble constructions, where neutral networks were used to learn a projection with more emphasis on previously misclassified instances similarly as in Boosting. Daoqiang Zhang [8] proposed constraint projections for ensemble learning, which use pairwise constrains to project original instances to a new data representation, then build base classifiers on the new data representation.

In this paper, considering the two conditions of ensemble and inspired by the idea in [8], we propose a novel method for constructing accurate and diverse classifiers through sampling pairwise constraints and subset selection based on samples clustering degree.

III. PROPOSED METHOD

A. Pairwise Constraints

The pairwise constraints[8] specify whether a pair of instances belong to the same class (must-link constraints) or not (cannot-link constraints). Such kinds of constraints have been widely used in several fields of machine learning, such as semi-supervised clustering. Pairwise constraints can be generated from class labels in advance. Given n labeled instances, we can derive approximately \(n^2\) pairwise constraints. Sampling pairwise constraint may help the base learners to have higher diversity because for n instances, there are at most \(2^n\) different results for sampling instances, but at most \(2^{n^2}\) different results for sampling constraints.
B. Subset Selection

Considering accuracy of the base classifiers, we adopt a subset selection scheme. For instances, suppose we have \( R \) subsets based on sampling pairwise constraints. Each subset \( S_i (i = 1, 2, ..., R) \) contains two sets: cannot-link set \( C_i \) and must-link set \( M_i \). 

\[ 
C_i = \{ (x_i, y_j), (x_j, y_i) \mid \text{belong to the same class}\}, \quad M_i = \{ (x_i, y_j), (x_j, y_i) \mid \text{belong to different classes}\}. \]

A distance function is defined to evaluate each subset’s clustering degree. \( n_a \) and \( n_c \) are the number of pairwise constraints in \( M \) and \( C \).

\[
D_i = \frac{1}{2n_{a_i}} \sum_{(i\neq j)\in M_i} (x_i - x_j)^2 - \frac{1}{2n_{c_i}} \sum_{(i\neq j)\in C_i} (x_i - x_j)^2 \tag{1}
\]

The first part of Equation (1) is the average distance among samples involved by \( M \) in each subset, while the last part of Equation (1) represents the average distance among samples involved by \( C \) in each subset. If the value of the first part is larger and the value of the last part is smaller, the statistical average distance \( D_i \) described in (1) is larger. Larger statistical average distance means the larger borders for classification. Therefore base classifiers trained with larger statistical average distance subset usually tend to have more accurate classification. Based on this idea, we propose the subset selection scheme. In the subsets selection, we select the subsets with larger value of Equation (1) from the sampled \( R \) subsets. The detail process of our algorithm is summarized in Algorithm 1.

C. A novel algorithm based on pairwise constraints and subsets selection

Algorithm 1:

Input: \( n \) training samples: \( \{ (x_1, y_1), ..., (x_n, y_n) \} \)

base learning algorithm: \( L \)

ensemble size : \( m \)

Process:

- a) Randomly draw a pairs of samples \( (x_i, x_j) \) from the original dataset with replacement.
- b) If \((y_i = y_j \text{ and } |M| < n_a)\), add \((x_i, x_j)\) into \( M \).
  
  Else if \((y_i \neq y_j \text{ and } |C| < n_c)\), add \((x_i, x_j)\) into \( C \).
- c) Repeat step 2 and 3 until \(|M| = n_a \text{ and } |C| = n_c\).
- d) Repeat step 1, 2, 3 until \( R \) subsets \( \{ x_1, x_2, ..., x_R \} \) are sampled.
- e) Compute \( D_i (i = 1, 2, ..., R) \) for each subset and sort \( R \) subsets according to their corresponding value of \( D \) and select the first \( m \) subsets \( \{ x_1, x_2, ..., x_m \} \).
- f) Train base classifiers: \( h_i (i = 1, 2, ..., m) \) based on the selected \( m \) subsets.

Output: the final hypothesis:

\[
H_s (x) = \arg \max_{h_i \in \mathcal{H}} \sum_{i \in \mathcal{H}} \left( x \right) \tag{2}
\]

From algorithm 1, we can see that our algorithm proceeds as follows: Given a set of labeled data \( \{ x_i, y_i \} (i = 1, n) \), and assume that we have obtained \( R \) subsets. For each subset, there are a set of pairwise must-link constraints \( S \) and a set of pairwise cannot-link constraints \( M \). Then we use equation(1) to compute the distance value of each subset, order the subsets according to their corresponding values and select the first \( m \) subsets in the ordering list. After obtained the subsets used to train base classifiers, what we should to do next step is to train base classifiers for the ensemble.

IV. EXPERIMENTS

In order to access the performance of our method described in the previous section, experiments are carried out on 6 UCI datasets in matlab7.0. C4.5 is adopted as our base learning algorithm and 10-fold cross validation is used to evaluate the error rate of classification. In experiments, assuming \( n \) instances for original dataset, \( n/4 \) pairwise constraints are sampled for \( C \) and \( M \) for each subset. In each subset, we can get \( n \) instances with replacement. Finally, the ensemble size is set to 5 for all ensemble methods. The main characteristics of 7 datasets are shown in Table I.

We compare our method with Bagging under a range of \( R \). Details about misclassification rate are shown in Table I. We can see from Table I that our method performs better that bagging in most cases.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Instances</th>
<th>Attributes</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glass</td>
<td>214</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Heart</td>
<td>270</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>Iris</td>
<td>150</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Segment(500)</td>
<td>500</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>Balance_scale</td>
<td>625</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Wine</td>
<td>178</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>Zoo</td>
<td>101</td>
<td>16</td>
<td>7</td>
</tr>
</tbody>
</table>

V. CONCLUSIONS AND FUTURE WORKS

In this paper, we present a novel approach for ensemble classifiers construction with pairwise constraints and subset selection techniques. Through experiments, it was proved that our proposed method demonstrated superior performance than bagging under the same environment. In the future, several issues can be studied more. First we only compare our method with bagging in this paper, we can compare our method with more existing popular ensemble methods. Secondly, how to evaluate the samples clustering degree for each subset can be studied more. Whether there exist the other optimal functions to evaluate it. Finally analyzing the proposed method theoretically is also an important direction.

ACKNOWLEDGEMENT

This work is funded by Scientific and Technological Project of Shandong Province (No. 2009GG10001003) and Independent Innovation and Achievements
Transformation Project of Shandong Province (No. 2009ZHZX1C0203).

REFERENCES

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Bagging</th>
<th>R=10</th>
<th>R=20</th>
<th>R=30</th>
<th>R=40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glass</td>
<td>0.28667</td>
<td>0.24571</td>
<td>0.22095</td>
<td>0.21905</td>
<td>0.19048</td>
</tr>
<tr>
<td>Heart</td>
<td>0.22770</td>
<td>0.20222</td>
<td>0.17260</td>
<td>0.17480</td>
<td>0.17420</td>
</tr>
<tr>
<td>Iris</td>
<td>0.04667</td>
<td>0.03867</td>
<td>0.06534</td>
<td>0.07200</td>
<td>0.00400</td>
</tr>
<tr>
<td>Segment (500)</td>
<td>0.05840</td>
<td>0.04360</td>
<td>0.03760</td>
<td>0.05200</td>
<td>0.03520</td>
</tr>
<tr>
<td>Balance scale</td>
<td>0.27742</td>
<td>0.19806</td>
<td>0.20871</td>
<td>0.18129</td>
<td>0.19387</td>
</tr>
<tr>
<td>Wine</td>
<td>0.02235</td>
<td>0.03294</td>
<td>0.01412</td>
<td>0.014112</td>
<td>0.02520</td>
</tr>
<tr>
<td>Zoo</td>
<td>0.74400</td>
<td>0.38750</td>
<td>0.43800</td>
<td>0.43400</td>
<td>0.56400</td>
</tr>
</tbody>
</table>