Comparing combiner systems using diversity measures

Husain Qasem
Telecommunication and Navigation Higher Institute
PAAET
POBox 4575, Alsalmia, 22046, Kuwait
hdashty@yahoo.com

Abstract: - We investigate the diversity of classifiers combined using different combiner systems. Three diversity measures are used to calculate the diversity of combined classifiers. We aim to find which combiner design method yields most diverse classifiers. We also aim to find if diversity measures are good indicators of system performances. The system combiner types are; Bagging and a conventional three classifier system, in which three classifier types are used; backpropagation neural network, bayesian and k-nearest neighbor classifiers. Results obtained on real data indicate the system with the higher performance yields more diverse classifiers, therefore, diversity measure is related to the system performance. We also found that bagging yields more diverse classifiers if neural network classifiers are used. However, the mixed classifier system yields more diverse classifiers if k-NN or bayes classifiers are used. This was in line with system performances.

Key-Words: - Combining, Fusion, Diversity, bagging, k-Nearest neighbor, MProduct

1 Introduction
The diversity [8] of component classifiers in a combiner system has been shown to yield better combiner performance. We aim to find which combiner system design method among the three investigated in this paper yields more diverse classifiers.

We experiment with three combiner types: Bagging [9], a conventional system consisting of three independently designed classifiers and a feature based combiner system proposed by Alkoot [13,14]. We repeat the experiments for each of the three types of classifiers used, namely; gaussian, backpropagation neural network and k-nearest neighbor classifiers. We calculate the diversity between the component classifiers in each combiner system using each of the classifier types. Three diversity measures were used to analyze the performance of each system. Results show that when the compared systems differ widely in performance, the diversity measures show the better system to be more Diverse.

2 Experimental Methodology
We experiment with different combiner systems where in each system we measure the diversity of each classifier. These include bagging and conventional systems containing multiple classifier types. We experiment with the k-NN, Neural network and gaussian classifiers. This comparison will serve our interest in knowing which combiner system yields more diverse classifiers. We experiment with real data from the UCI [11] repository. The data sets used are BCW, Ionosphere, Wine, Glass, car and Lung.

2.1 The MProduct Fusion method
The fusion method used in the multiple classifier systems is the Modified Product, [12], MProduct. The reason for choosing it is that Product is closely related to the Bayesian rules it was derived from, while sum is derived under restrictive assumptions. MProduct has shown a great improvement over Product and on average it outperforms Sum. MProduct is a fusion method of combining classifier outputs that alleviates the veto effect. The veto effect in the Product fusion is caused by small classifier measurement output values dominating the product, i.e. giving a close to zero result. This commonly occurs when noise levels are high. The basic idea is to modify the output of an expert if it falls below a specified threshold. For all experts, we examine the output for one class at a time. Hence for an expert its estimate of the probability of one class may change while its estimates for other classes may stay the same.

2.2 Diversity Measures
To compare the diversity between two classifiers assume the probabilities for their respective pair of classes are $P_1$ and $P_2$. The diversity of classifiers $D$ can be calculated as:

$$ D = 1 - \frac{1}{2} \left( \frac{P_1 + P_2}{2} \right)^2 $$
correct/incorrect outputs over the full test set are as shown next:

<table>
<thead>
<tr>
<th></th>
<th>Dj correct</th>
<th>Dj wrong</th>
</tr>
</thead>
<tbody>
<tr>
<td>Di correct</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Di wrong</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

Then we can calculate the diversity [8] between the two classifiers using any of the following three measures:

a) The disagreement measure:
   \[ D_{i,j} = b + c \]
   This measure represents the probability that the two classifiers will disagree on their decisions. The closer it is to 1 the more diverse the classifiers are. This measure is simple and very representative of the relation of the two classifiers. In the tables of diversity this measure is referred to as \( \text{disa}_{ij} \) where \( ij \) stands for the compared classifiers \( i \) and \( j \).

b) The correlation measure:
   \[ \rho_{i,j} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \]
   This is the well known correlation measure. It is represented by \( r_{ij} \) in the tables of diversity.

c) The Q statistic: Yule’s Q statistic for classifiers \( Di \) and \( Dj \) is:
   \[ Q_{i,j} = \frac{(ad - bc)}{(ad + bc)} \]
   It is related to the distance measure, and finds the normalized difference between the agreement and disagreement of the two classifiers. The closer it is to zero the more independent the classifiers are. When it is positive it indicates that the two classifiers tend to agree on their decisions.

3 Experimental Results

Classifier performance is related to diversity of component classifiers. We will see results in the form of combiner performances, followed by an analysis of the diversity results for the different classifiers and systems.

Experimental results show that mixed classifier system outperforms bagging when using kNN or bayes classifiers. However bagging outperforms the mixed classifier system when a neural classifier is used. This is in line with Briemans finding that bagging does not perform well on stable classifiers like kNN. Therefore, we expect the diversity of neural classifiers to be larger than kNN. Also, since bayes classifiers yield lower classification rates we expect its diversity to be lower than kNN.

As indicated in section 2.2 we calculate the diversity between pairs of classifiers for each classifier type and for all data sets and combiner systems. The diversity measures show the difference between classifiers built using each combiner system.

<table>
<thead>
<tr>
<th></th>
<th>Wine</th>
<th>Iono.</th>
<th>BCW</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mixed classifier system</strong> (kNN, neural network, bayes)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validation set rate</td>
<td>99.25</td>
<td>96.21</td>
<td>96.78</td>
</tr>
<tr>
<td>Test set rate</td>
<td>94.73</td>
<td>90.36</td>
<td>96.22</td>
</tr>
<tr>
<td><strong>Bayesian classifier bagging</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validation set rate</td>
<td>32.95</td>
<td>-</td>
<td>90.55</td>
</tr>
<tr>
<td>Test set rate</td>
<td>92.43</td>
<td>-</td>
<td>92.86</td>
</tr>
<tr>
<td><strong>k-NN classifier using bagging</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validation set rate</td>
<td>93.32</td>
<td>77.57</td>
<td>96.14</td>
</tr>
<tr>
<td>Test set rate</td>
<td>94.43</td>
<td>80.99</td>
<td>95.78</td>
</tr>
<tr>
<td><strong>neural network classifier using bagging</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Validation set rate</td>
<td>95.66</td>
<td>89.74</td>
<td>96.06</td>
</tr>
<tr>
<td>Test set rate</td>
<td>97.57</td>
<td>92.39</td>
<td>96.02</td>
</tr>
</tbody>
</table>

Table 1 classification rate for the mixed classifier combiner and the bagging combiner using different classifiers

Comparing the mixed classifier system to the bagging, we find that when neural network classifiers are used bagging system yields more diverse classifiers than the mixed classifier system. The opposite is true when using k-NN or bayes classifiers, where the mixed classifier system yields more diverse classifiers than bagging. These diversity results are similar to the performance results as indicated in table 1. For all data sets bagging using neural network outperforms the mixed system, while bagging using k-NN or bayes classifiers underperforms the mixed classifier system.

The Q statistic is more related to the performance of classifiers showing bagged neural classifier with the lowest Q value for all data sets. The k-NN classifier outperforms bayes for all data sets, however for
Wine the bayes Q statistic is lower than that of a k-NN, indicating more diversity than k-NN.
Looking at the correlation measure we find that the neural network classifier yields the least correlation between classifiers, indicating higher diversity. However, for the Wine and Ionosphere the correlation of bayes classifiers is lower than k-NN. In general for bagging we find all the diversity measures indicate that the best classifier, i.e. neural classifier, is more diverse. However, bayes classifier which yielded a lower classification rate than k-NN was found to be more diverse. This indicates bayes classifiers commit more errors but perform differently, while k-NN make more accurate decisions and commit similar error, hence less diverse.

4. Conclusion
In our experiments we compared the diversity of classifiers built using three different models in two combining scenarios. The different classifiers used were k-NN, Normal and a three layer neural network perceptron. The combining systems used were bagging and a conventional mixed classifier system which includes one of each of the three classifier types. We used three diversity measures, the distance, the Q-statistic and the correlation measures. We measure the diversity between classifiers designed by the different methods to find if they correspond proportionally with the system performance, and if so which system yields more diverse classifiers.
We found that the diversity measure correspond with
the system performance. The Q-statistic was the measure that had the closest correspondence between the performance and diversity. The Q-statistic results show that the best classifier, i.e. neural network, yields the highest diversity results while the bayes classifier yields the lowest. The Q measure has more information than the distance measure. We also found that bagging yields more diverse classifiers when neural network classifiers are used, while the mixed classifier system yields more diverse classifiers when k-NN or bayes classifiers are used.

References:


