Ensembles of classifiers are among the strongest classifiers in most data mining applications. Bagging ensembles exploit the instability of base-classifiers by training them on different bootstrap replicates. It has been shown that Bagging instable classifiers, such as decision trees, yield generally good results, whereas bagging stable classifiers, such as Naïve Bayes, makes little difference. However, recent work suggests that this assumption does not apply to the Data Stream Setting.

Bagging

Bagging [4] is a technique that trains various models on different samples of the data and combines the predictions.

Algorithm Selection on Data Streams

Continuation of the Meta-Learning Experiment presented in [2].

Experimental Results

The experiment contains two classifiers (Naïve Bayes and $k$-NN, with $k = 10$) and two Bagging Schema’s (Online Bagging and Leveraging Bagging). The images show the results of both schemes on Naïve Bayes.

Statistical Tests

Two statistical tests were performed:

- T-Test found no significant differences.
- Wilcoxon Signed-Ranks test found significant difference in many cases.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Online Bag</th>
<th>Lev. Bag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td>no</td>
<td>Lev. Bag</td>
</tr>
<tr>
<td>$k$-NN</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table 1: Wilcoxon Signed-Ranks Test results, 95% confidence.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Online Bag</th>
<th>Lev. Bag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naïve Bayes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>$k$-NN</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

Table 2: T-Test results, 95% confidence.

Conclusions

A possible explanation can be found in the fact that the Wilcoxon Signed-Ranks Test bases its conclusion on the signs of a classifier; it only considers whether one schema was better, equal or worse on a given data stream. The T-Test bases its conclusion on actual scores. The fact that the Wilcoxon test found statistical evidence that bagging actually improves the performance of stable classifiers in the data stream setting, but the T-Test not, leads to the belief that improvements can be obtained, but these are very limited. More research is required to give a decisive answer to the question whether Bagging Stable Classifiers works on Data Streams.

References