

Result

Data used in classifier

| | Vbulletin | PhpBB | I.P. Board | SMF | Burning board |
|---------------|-----------|-------|------------|-----|---------------|
| Training data | 100 | 70 | 40 | 20 | 5 |
| Unclassified | 40 | 40 | 30 | 12 | 8 |

Table 1: Data used in classifier

Table1 shows downloaded data we used in training and classification. From the table we derive that we have a lot of data templates from “Vbulletin” and “phpBB” boards. These two are very commonly used as board engines, and was therefore easy to locate templates from. “Invision Power Board” and “SMF” is not widely used as board engines, so we downloaded a few easy found examples. These examples is not a quite representative selection of the different board versions, but could give a indication of the general accuracy of the classifier. Burning board is a even more seldom found board type. We therefore decided only to include it in our data collection, because it have quite distinct features in both classes and id attributes etc. This could also be very useful when testing how the classifier would react, whit only a small training set.

Confusion Matrix

| | SMF | Invision Power Board | Vbulletin | phpBB | Burning board |
|----------------------|-----|----------------------|-----------|-------|---------------|
| SMF | 12 | 0 | 0 | 0 | 0 |
| Invision Power Board | 1 | 28 | 1 | 0 | 0 |
| Vbulletin | 0 | 0 | 40 | 0 | 0 |
| phpBB | 0 | 0 | 4 | 36 | 0 |
| Burning board | 0 | 0 | 1 | 0 | 7 |

Table 2: Confusion matrix

Table 2 shows our confusion matrix that displays the test results. The vertical column is the actual board type, and the vertical rows is what the classifier classifies it as. From the confusion matrix it is clear that all the boards from “Vbulletin” and “SMF” are classified correct.

Accuracy

| Board type | Accuracy |
|----------------------|----------|
| SMF | 1 |
| Invision Power Board | 0.93333 |
| Vbulletin | 1 |
| phpBB | 0.9 |
| Burning board | 0.875 |
| | |
| Total: | 0.9417 |

Table3:

Table3 shows the accuracy of the different boards and the total accuracy of the classifier. The accuracy for each board type is the number of classified boards divided by the total number of that board type in the classification set.

The accuracy of the “phpBB” board type is quite low because some of the testing examples are quite modified from the original template. This would possibly classify better, using more training examples from the specified board type.

The accuracy for “Burning board” is the lowest accuracy in table3. The reason for this is, as we can see from table1, it is only used 5 boards to train with and 7 to classify. This is a bit unbalanced but it is still 7 out of 8 (see table 2) templates classified correct.

The total accuracy of the classifier is the sum of each discussion board accuracy divide by the total of boards tested.