Decision Tree Learning Experiments: Analysis and Conclusion on Tic-Tac-Toe

marbille.juntado

May 2017

1 Results

1.1 Experiment 1

Performing the experiment and building a decision tree five times lead to relatively high accuracy results. For each iteration, there proved to be minimal difference between each corresponding accuracies.

The accuracy of the experiment is computed as follows:

$$Accuracy = \frac{No.\ of\ Correct\ Predictions}{Total\ No.\ of\ Test\ Cases}$$

The predicted classifications are then mapped in a confusion matrix (*vis-a-vis*) the actual classifications where each cell contains the average count (sum of the iterations divided by five) of the observed classifications from the experiment.

The first experiment produced the following confusion matrix.

Confusion Matrix		
	Predicted (Yes)	Predicted (No)
Actual (Yes)	270.2	33.6
Actual (No)	41.0	119.2

which resulted to the following accuracy:

$$Accuracy = \frac{270.2 + 119.2}{479} = \frac{389.4}{479} \approx 0.81$$

An accuracy of 0.81 means about 4 out of 5 test cases, the decision tree will predict the outcome correctly.

1.2 Experiment 2

In order to perform the second experiment, 20 training examples have been randomly selected and noise has been introduced by altering the classifications to "Maybe" since this word has not been identified by the program as a negative or positive classification.

The second experiment produced the following confusion matrix.

Confusion Matrix		
	Predicted (Yes)	Predicted (No)
Actual (Yes)	253.6	42.0
Actual (No)	35.2	121.2

which resulted to the following accuracy:

$$Accuracy = \frac{253.6 + 121.2}{479} = \frac{374.8}{479} \approx 0.78$$

2 Conclusion

There is a slight bump down between the accuracy results of the first and second experiments. From the difference of about 2% between the results, we can posit that the decrease in accuracy may be proportional, if not equal to the amount of noise in the training data set. Since,

$$\%\ Noise = \frac{no.\ of\ noisy\ examples}{total\ no.\ of\ training\ examples} = \frac{20}{479} \approx 0.04$$

Although, it is a minimal difference for this particular data set, we have to take into account that the accuracy may suffer a bigger loss if noise-to-training examples ratio is increased. This is an example of the limitations of decision tree learning. What can be done to improve decision tree learning algorithms (especially ID3) is by autonomous backtracking, information gain reduction, and other ways to overcome the disadvantages of ID3.