Kendra Maggiore

CS 4315

2/15/2020

Assignment 1

I compared the Random Tree algorithm to the J48 unpruned algorithm using Weka, and ten UCI data sets. The J48 algorithm is the open source Java implementation of C4.5. To do the experiment, I would load the data set into Weka then run both algorithms on Weka and record the results. For the average tree height, I used the visualize tree option in Weka and counted the trees layer by layer. For the Random Tree algorithm, I left the setting as they were set in Weka. When using the J48 algorithm, I went into the options, and checked the box to make it unpruned. I also made sure before running the algorithms that it was set to the class attribute in the drop-down menu on Weka.

The J48 unpruned algorithm produced very good result looking at **figure 1**. The accuracy results were on average about 79 percent with it only performing under 70 percent on that glass data set. It was even able to achieve 100 percent on two different data sets the Autism-Adolescent data sets and the mushroom data set. Its average tree height was also short averaging about 8 layers per data set.

The Random Tree algorithm produced average results looking at **figure 1**. The accuracy on average for Random Tree was 72 percent with the lowest score being 62 percent on the zoo data set. It only achieved 100 percent on the mushroom data set. Random Tree’s average tree height was about 9 layers per data set.

When comparing these two algorithms on accuracy on average, J48 unpruned was more accurate then Random tree. There were a few instances where they were very close (within a few percentages of each other), and one time on the mushroom data set where they both achieved 100 percent accuracy. There was only one instance on the glass data set that Random tree was more accurate then J48 unpruned.

J48 unpruned also had an average tree height that was shorter than Random Tree. There was one instance on the audiology data set where they both had the same tree height of 14. J48 unpruned produced a better accuracy in this instance. There were also two instances were Random Tree was shorter than J48 unpruned. This occurred on the zoo data set, and the anneal data set. Again, J48 produced better accuracy results on these datasets even though it had a longer tree. This leads me to believe that a shorter tree length does not necessarily mean the algorithm will produce better accuracy results in all cases.

In conclusion, the J48 unpruned algorithm is an all-around better algorithm when compared to Random Tree on the data sets that I tested with. J48 unpruned was able to produce better accuracy results while also having a shorter tree on average. The Random Tree algorithm did an acceptable job with the data sets but could not keep up with J48 unpruned.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Dataset | num of instances | number of attributes | number of classes | Random average height | Random average accuracy | J48 Unpruned Average height | J48 Unpruned accuracy |
| Autism-Adolescent | 104 | 21 | 2 | 3 | 80.7692 | 3 | 100 |
| anneal | 898 | 39 | 6 | 10 | 98.3296 | 13 | 98.441 |
| audiology | 226 | 70 | 24 | 14 | 65.4867 | 14 | 76.9912 |
| iris | 150 | 5 | 3 | 6 | 92 | 4 | 96 |
| zoo | 101 | 18 | 7 | 2 | 62.3762 | 6 | 92.0792 |
| mushroom | 8124 | 23 | 2 | 9 | 100 | 6 | 100 |
| balance-scale | 625 | 5 | 3 | 13 | 77.28 | 11 | 79.2 |
| glass | 214 | 10 | 7 | 14 | 70.0935 | 11 | 67.2897 |
| hepatitis | 155 | 20 | 2 | 13 | 76.7742 | 10 | 80.6452 |

**Figure 1**- Data Results from J48 unpruned and Random Tree