

COMP4331 Assignment 2 Report HUANG, Zijian 20413784

2.

2.1

- (1) The Environment: MacBook Pro 2018
Processor: 2.3 GHz Quad-Core Intel Core i5
Memory: 8 GB 2133 MHz LPDDR3
Graphics: Intel Iris Plus Graphics 655 1536 MB
OS: macOS Catalina (version 10.15.4)

(2) Running Time:

- ① ID3 Decision Tree:
- Train: 6.440241098403931 seconds
 - Test: 0.0020470619201660156 seconds
- ② C4.5 Decision Tree:
- Train: 10.387693166732788 seconds
 - Test: 0.0016567707061767578 seconds
- ③ Naive Bayes Classifier:
- Train: 963.0644071102142 seconds
 - Test: 0.003103971481323242 seconds

(3) Result

① ID3 Decision Tree:

not_recom
spec_prior
priority
very_recom
very_recom
priority
not_recom
spec_prior
very_recom
very_recom

② C4.5 Decision Tree:

not_recom
spec_prior
priority
very_recom
very_recom
priority
not_recom
spec_prior
very_recom
very_recom

③ Naive Bayes Classifier:

not_recom
spec_prior
priority
priority
very_recom
priority
not_recom
spec_prior
priority
very_recom

(4) Explanation

Firstly, for the running time, C4.5 Decision Tree's training time is a little bit longer than that of ID3 Decision Tree because several 'SplitInfo' need to be calculated in every split step. Naive Bayes Classifier's training time is much longer than that of Decision tree because it needs to compute all of the possible combinations of the data attribute in this assignment, although we do not need to compute like this when we apply this algorithm. The testing times are quite similar because they all only need to check in the model once for a new data point.

Secondly, considering the results, two kinds of decision tree have the same result because the algorithms are quite similar and the class distribution in the dataset and every node is relatively uniform for each class.

After I check the table between each pair of classes, I find that they have little relationship between each other. Also, after I print the decision tree out, I find that it does not use the 'form' and 'children' attributes, which means decision trees may do not consider enough. Therefore, I believe that the results should be the same with the Bayes' results.

(5) Results that I believe:

not_recom
spec_prior
priority
priority
very_recom
priority
not_recom
spec_prior
priority
very_recom