

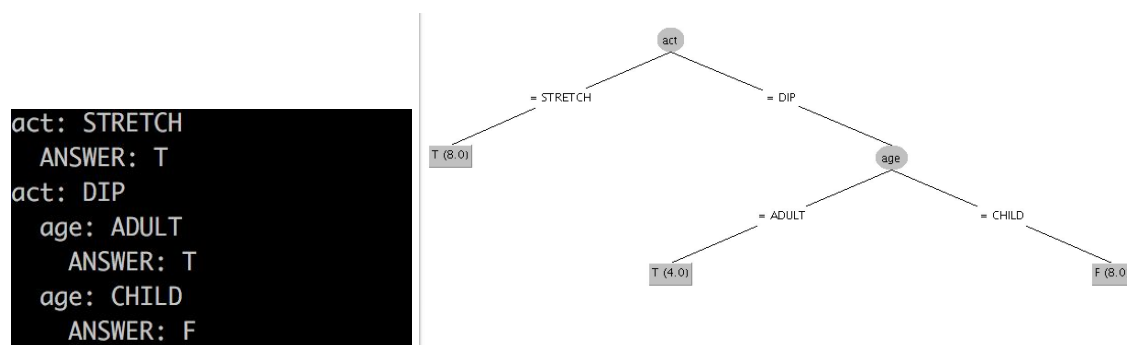
Decision Trees Report

The main advantages that can be obtained from writing a program to represent this decision trees algorithms like we did with id3, is maybe to be able to customize certain rules of how the tree should be built or being able to make modifications to existing algorithms for research purposes. Another parameter that should be taking into consideration is the fact that, some of the main disadvantages could be runtime or efficiency, as the developer not necessarily uses the appropriate data structures or algorithms to make some subroutines of the program like in the process of obtaining entropy, information gain or simply in the way the attribute was stored. In my opinion, using the algorithms that are contained in Weka would be better in this case, since I belief they try to optimize the software plugins the most they can.

A reason why I chose the balloons-actions dataset was because this break I started to go to the psychologist, this was a strange coincidence and I read a little bit of the paper in which this dataset was used. The experiment tried to explain if the previous knowledge can influence the rate of concepts learning, and I think this is connected to the content we saw on the first lessons because the learning that related the most was the one made in a logical approach of relating casual concepts to prior knowledge.

Some other reasons that made my life easier by using this dataset, was the fact that its attributes were discrete and none of them were missing in the actual data. I tried some other datasets and my program failed if I had missing attributes on the data.

The trees I generated with both Weka and my program resulted to be the same, I think this is mainly because there were not so much attributes in the dataset so the complexity did not grow a lot, so that there was no difference in the order of the attributes id3 or the improved version j48 made the split process. J48 maybe would produce a different tree if there were so much attributes or if in some part of the tree the branches could be pruned.



Since decision trees are used to classify some new input data from the model it produced I would use it in problems where I need to know the expected output of something given its characteristics values. I could use id3 to make predictions on nominal values or c4.5 with numerical values. One interesting application could be to do medical diagnosis to identify differences between some similar diseases or to decide if I should buy a car based on its specifications and my needs.

References:

Pazzani, M. (1991). The influence of prior knowledge on concept acquisition: Experimental and computational results. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 17, 3, 416-432.