Harlow Huber

Research Assistant

10/3/2022

GOALS

* Work on expert data mining program
  + ~~Go to successful chains after asking majority flag questions~~
  + Ask about if a particular attribute must be true
  + Allow user to order questions manually (show table in question)
    - Use excel
    - Tell user to put numbers in second column for the planned order
  + Dynamic - ~~If bottom vector in chain is class 0, then skip to another chain in hope that it has a class of 1~~. If top vector in chain is class 1, then skip to another chain in hope that is has a class of 0.
  + ~~Actual order -> query order~~
  + ~~Planned order -> fix for majority flag~~
  + ~~Updated order -> where planned order diverges from dynamic questioning~~
    - Update flag -> 1 where orders are diverged
* Presentation?
* Working on k-value Hansel Chain visualization

RESULTS

I have done all the crossed out goals in the above list, so now there is an Planned, Updated, and Final Query Order column in the output table. I have added the option to make a particular attribute always true, but I haven’t tested it yet.

Also, I added the dynamic option for the program that we discussed, but right now it works with just the bottom vector of a Hansel Chain: if that bottom vector is a 0, then we skip that chain and go onto the next, and the last step is to go over those vectors we skipped. However, the results aren’t that good for the 5D real estate dataset’s monotone Boolean function. I theorize that starting at the top of the Hansel Chains will yield better results for that particular function, but I’m not sure about functions in general. I would hatch a guess that it would also be better in general, provided that the user is picky with the values for the datapoints (I believe this to be likely). If most of the datapoints turn out to be true, however, then I believe starting at the top of the chain might yield worse results.

TO DO

* Work on expert data mining program
  + Ask about if a particular attribute must be true
  + Allow user to order questions manually (show table in question)
    - Use excel
    - Tell user to put numbers in second column for the planned order
  + Dynamic - If first vector in a long chain is class 0, then we may potentially go through all the elements in that chain to find the lower one, but first explore in all chains what is the value of the lowest element. If in one chain the lowest element is one, then it expands that chain and possibly the previous chains. Essentially, go to another chain in hope that it has a class of 1. Heuristic to minimize number of questions. If top vector in chain is class 1, then skip to another chain in hope that is has a class of 0.
  + Majority flag even when not used
  + ~~natural order -> default order~~
* presentation
* Working on k-value Hansel Chain visualization