Testing The Performance Of Minimal Models Trained On Sparse Data

Background

DFA learning is the problem of recreating a DFA from a collection of traces. Two popular techniques for this are DFASAT (optimal) and EDSM (heuristic). Both of these are in the FlexFringe software package. To develop training sets, sampling from DFAs is often done. Properties in the generating DFA can have downstream effects in the tests results of leared models.

Hypothesis

The sampling procedure used in the STAMINA competition was hypothesised to have interesting emergent properties. It performs a random walk, terminating in final states with probability 1/(1 + 2*deg(q)). Negative traces are modified from positive traces. Positive traces will be distributed similarly to a geometric distribution, and negtive traces binomially around positive traces. From this, we can hypothise the following:

1. Test performance is distributed binomially around the number of final states.

2. Increased alphabet size results in longer traces, and equivalent test performance.

3. Increased node count does not effect trace length, and results in lower test performance.

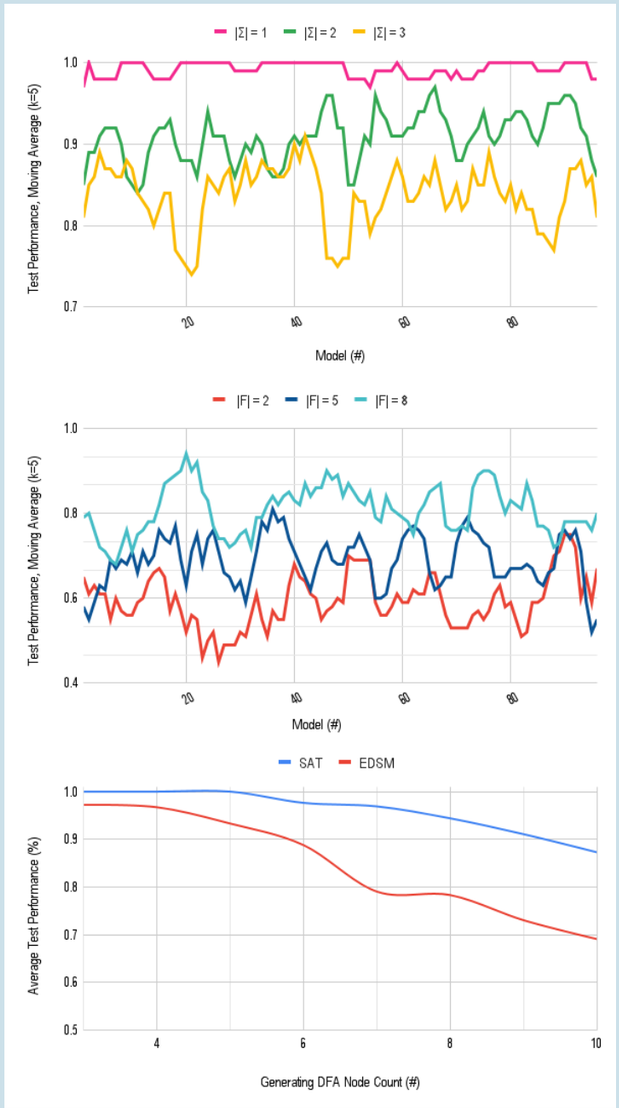
4. Optimally trained models will be more likely to have a better test performance than equivalent heuristically trained models.

Methodology

For training models, random DFAs were generated andthen sampled from using the STAMINA techique. To generate a DFA, a sparse tree is first generated. For each node, outgoing edges are then assigned. Finally, a number of states are randomly selected to be final.

For each hypothesis, a collection of data sets was created to test whether this hopythesis held. Statistical difference between groups of resulting data was tested for using a T-test. Models are learned using FlexFring and tested using 5-fold cross validation.

Results



Analysis

It was found that alphabet size resulted in significantly longer words. This still resulted in a lower test performance as alphabet size increased.

Increases in final state count resulted in longer traces when sampled. Surprisingly, a 20% and 50% proportion of final states did not result in a significantly different test performance. Models with an 80% proportion had a significantly higher test performance. This indicates that the effect of final state proportion on test performance is strongest at high final state counts.

Increased node count did not result in a difference in the distributionn of sampled traces. It resulted in significantly lower test performances in resultant models.

Models trained heuristically had consistently lower test performance than their optimally trained counterparts.

Conclusion

The experiment was ultimately succesful, as though some hypothesis did not resolve as expected, useful results were found. This could prove useful for future research that relies on similar sampling methods, or seeks to design their own DFA learning techniques.