Project Deliverable 3

Andrew Berry

October 10, 2016

I pledge my honor that I have neither given nor received any unauthorized aid on this work.

1 Clarifications from Deliverable 2

No additional details were requested in the previous project update.

2 Accomplishments and Scheduling

Since the submission of the previous deliverable, I have assembled a working non-distributed version of the random forest algorithm, which will be adapted to a multi-agent system for the final result. The non-distributed version of the program was able to correctly predict the state around 10% of the time, while the correct "region" was chosen between 35 and 40% of the time. These results seem promising because random guessing would produce a correct result approximately 2% of the time for states, and approximately 16.67% of the time for regions (if we break the country into six different regions). I am now working on a simple multiagent classifier to run against my dummy dataset as a proof-of-concept to segue into developing the final multiagent classifier for the hospital dataset. This first pass of the dummy dataset is expected to be completed by the end of this week. Project progress is roughly in line with the initial schedule, with no major anticipated roadblocks.

3 Knowledge and Beliefs

All agents in my system will share the same initial knowledge base, which will be the training set from the hospital readmission and death rate dataset. If it determined later that the multi-agent system needs the added flexibility of having the training sets differ from agent to agent, that can be implemented as well, to give them each different knowledge.

Degrees of belief factor into my system in that any agent after the first one will consider the "vote" of the agent that precedes it as an alternative to the "vote" it finds of its own accord. The likelihood with which the agent chooses its predecessor's answer over its own independent answer is based on the number of estimators used by each of the two agents. Suppose that agent 1 has voted "AZ" and came to this vote using a random forest with 10 estimators. Suppose also that agent 2 finds that its random forest suggests "MO" using 5 estimators. If we take the list of these two values, $\langle 10, 5 \rangle$, and normalize them to have them represent probabilities, we get $\langle 0.67, 0.33 \rangle$, meaning that agent 2 is twice as likely to choose "AZ" as "MO" because it has a higher degree of belief about agent 1's choice, or more "faith" in agent 1's ability to produce a correct result (although it may not always be the case that the agent with more estimators correctly predicts the answer).