Milestone Report

My project involves investigating and implementing some sort of accuracy measure within neural semantic parsing tasks. Specifically, I am targeting the task of code generation, in which a natural language description is given, and the model outputs code of some language to perform that description. So within this context, the accuracy of the model is what percentage of the test set it outputs correct code for. I originally identified a few sources for learning various methods of incorporating accuracy in the training and decoding portions, and decided upon a milestone that would involve learning what state of the art models have attempted and the benefits and drawbacks of each method.

I set the task of reading through these papers, getting a general background understanding, and deciding upon which methods to implement first as the milestone for the end of this semester. In this respect, I have met the milestone described in my original proposal. I read through Norouzi et al. [3], as well as Keneshloo et al. [1], which detailed current work in combining reinforcement learning techniques with sequence to sequence models. It seems like the original technique I proposed in the talk, which is policy gradient, is the most easy to implement method. While it is easy to implement, it is also has very high variance and won’t provide any guarantees as to an improvement in performance. Regardless, it seems like a good first model since it is both achievable and theoretically better.

Because my first milestone was something pretty simple to achieve, namely gaining a better background understanding and deciding which method seems most viable, there have been no major surprises thus far. As far as changes to the milestone and overall project proposal, it might be in my interest to spend some more time thinking about the task, perhaps focusing on a more well established semantic parsing task. This is because the task of code generation still doesn’t have many great metrics, mainly using BLEU and accuracy, whereas some other tasks might provide more metrics to be used in the RL component.

Even though I plan on spending some time re-assessing if code generation is the best task for this, it should not impact the milestones I have set for 15-400. If anything, a switch away from code generation may make the milestones easier to achieve as the switch will be to a more established task. Lastly, as far as resources to complete this project, I have everything that I need. All of the papers I cited in my proposal are open source with sources on github, and I have acquired an environment in which I can train and test the models.
References


