Jorge Arellano

Professor Doyle

Ling 354

2/8/18

Botnik passage

*For some monsters, the game plan in fact has been skills of stress. When you've just nuked things slowly, British monsters are dangerous and they feel like powerful dragons. The only solution for immigrants can also asphyxiate invaders from every treasure. They probably want something secure from shopkeepers like Polaroids, if possible to avoid human fatalities.*

The passage I chose to write had many surprises that I learned about the algorithm, as well as myself. The beginning of the piece was chosen with the word “for” because it is a good starting position for most story-oriented writing. What I found was at best a ~2-word suggestion generation that caused the passage to strangle itself with very few choices. My goal was to write a coherent passage that would make sense for anyone to read irrespective of reading level/age. What I found was something far more complex and time consuming as well as incoherent.

The nature of the passage is a preface to a dark story about monsters and mystical animals. The outcome was an incoherent attempt to do just that. I noticed that after using a “big word” like monsters, many “smaller” > 5 letter words were generated for the next word. This was a problem because they most of the worlds that were generated were mostly useless. The word “Africa” was generated after “Monsters” which would have been a non-sequitur if used. Although my task was to make the passage coherent, I noticed that the longer the sentence was, the less likely my goal was going to be. I only used 1 comma per sentence because this allowed me to somewhat control the outcome.

The biggest problem was the bottleneck, if let’s say in any given sentence the next word I were to write had 15 correct combinations that could be used, only 2 were at my disposal with Botnik. If we use the 2-word maximum, we in some sense would be only using 10% of the next-word choices. I believe the reason the generator does this is probably because it is hard coded to rank the amount of times the next-word is used from the training data it received. So, in some sense there is a bias towards using words that are necessary for the prose it was trained on, but do not generalize in other contexts.

After finally facing the fact that the passage I was going to write, was impossible I started to add random words that would be sufficient to complete a sentence, but make the previous and post sentences ridiculous. This was especially fun when I stumbled onto the word immigrant, which is completely random if someone is trying to write about a mythical story. The grammar that followed was just weird, but it did have meaning in so far as it would make sense in the English language. In conclusion, I think the Botnik is close to being able to make coherent choices of what next-word likelihood should be, however it is very far at the same time. The word bottleneck is probably the biggest problem, or maybe it is the human-in-the-loop aspect which causes all of these issues.