CS 4740 – NLP

Team Members: Michas Szacillo (mas744), Alexandra Ward (amw349), Brandon Sweezer(bas329), and Abiy Tibebu (agt33)

Project 3 Proposal

Baseline System:

For the baseline, we decided to simply use a Random Guess model. To do this, instead of simply returning a random word from the given paragraph, we decided to use IOB tags to make the list of possible words to return smaller and more accurate. To do the IOB classification, we decided to use the Stanford NER module.

Our decision to use random guess is mostly so that we have an extremely simple model to start from, but to also start gaining intuition on the QA problem itself. However, we did go a step further than returning just a random word or phrase, since we are classifying the paragraph contents by IOB tag.

Baseline Results:

|  |  |
| --- | --- |
| Exact Match | F1-Score |
|  |  |

Proposal:

To begin, we plan on trying some more simple models such as the Window Sliding model used in the critique article. We also plan on playing around with tagging systems to make picking out the answers in the paragraph easier (for example POS tags, or IOG tags). As we continue to improve our model, we will look into possible linguistic NP filters as well, since we know that most of the answers to the questions are phrases and not just words.

Similarly, we plan on classifying question types and semantic class types so that we can narrow down the possible answers even further to a corresponding return type (e.g. Person, Place, etc).

Neural networks?

After we have found an answer, we plan to evaluate the validity of that answer using a method we will call statement rearrangement perplexity (SRP). SRP will rearrange the phrasing of the question and the top three answer possibilities we have highlighted so that they form a statement. Then it will evaluate the perplexity of each of these three sentences, using N-gram (we will only do bigrams as trigrams often don’t add much, as learned in previous assignments) models generated on the paragraph and question. The least perplex answer will be selected as our final answer. We believe this will improve our accuracy, for the correct answer will have extremely low perplexity, as the rearranged statement will have almost identical phrasing to that of the paragraph.

Group Contribution:

Brandon:

Alexandra:

Michas:

Abiy: