Group: Shijing Zhong(sz6539), Michael Li (mal4565)

Suggestions: http://www.cs.utexas.edu/~xcliu/Project_Suggestions.pdf
Papers:

1) https://arxiv.org/pdf/2002.12327.pdf (survey)
    a) BERT encodes syntactic structure but doesn't necessarily use it when predicting
    b) Even when attention heads specialize in tracking semantic relations, they do not necessarily contribute to BERT's performance on relevant tasks. https://www.aclweb.org/anthology/D19-1445/
    c)
2) https://arxiv.org/pdf/1706.03762.pdf (attention is all you need)
    a) Individual attention heads exhibits behavior related to the syntactic and semantic structure of the sentences,
3) https://ai.tencent.com/ailab/nlp/papers/emnlp2018_deep_representations.pdf (deep representation for NMT)
    a) Tried different things to connect information/data from lower layers to upper layers
4) https://arxiv.org/pdf/1807.03819.pdf (Universal Transformers)
    a) Recurrent in the # of changes for each embedding/transformer weights
    b) "UT does not recur over positions in the sequence, but over consecutive revisions of the vector representations of each position (i.e., over "depth")"
5) https://arxiv.org/pdf/1810.04805.pdf (BERT)

6) https://arxiv.org/pdf/1905.10650.pdf(Are Sixteen Heads Really Better than One?) https://www.aclweb.org/anthology/D19-1445.pdf(same attention patterns, they have varying impact on performance across different tasks...)
7) https://arxiv.org/pdf/1906.01698.pdf
8)

Blogs:
https://nlp.stanford.edu/~johnhew/structural-probe.html

1. **statement of problem:** Something something NLP/Transformers
    a. There are plenty of papers noted on the existence of syntactic structure inside BERT embedding, but we don't know whether the model is using that knowledge.
    b. Given that previous papers talked about the existence of syntax within BERT, we wanted to quantify how much of that syntax knowledge BERT uses to make its decisions.
2. **what you intend to do:**
    a. Experiment over the syntactic structure and study its correlation to the prediction accuracy. Compare low and high accuracy prediction samples, and examine its syntactic structures' difference

b. How much does BERT use syntax/the encoded syntax tree in its predictions?
    i. Shuffle the words randomly (has been done in previous work)
    ii. We modify an approach from Stanford that derives the syntax tree from the embedding space such that we will perturb the embeddings such that the projected space with the syntax tree will be different. Then we will pass this through BERT and evaluate the performance differences.
    iii. Use the syntax tree (perhaps from the Stanford Parser) and shuffle different Noun Phrases or verb phrases or something to perturb the sentence
    iv. Design a experiment that checks if BERT relies on the syntax knowledge encode in the embedding
c. (if time permits) Apply these findings and see if this is preserved when BERT is compressed
    i. Studies have shown that you can reduce the total amount of heads and still maintain the same amount performance
    ii. Discover a way to extract the syntactic structure in the embedding of BERT and analyse its importance in the inference stage.

3. **what data you will use:**
    a. https://huggingface.co/transformers/model_doc/bert.html
    b. https://github.com/john-hewitt/structural-probes (deriving syntax tree from embeddings)
    c. We are examining the BERT-Base Embedding and probably BERT-Large Embedding based on hardware and time constraint.
    d. Penn Treebank for syntactic parsing
    e. https://nlp.stanford.edu/software/lex-parser.shtml
4. **a rough timeline:**
    a. Set up the dataset and models
        i. Setup perturbations similar to structural-probes
            1. Setup module to perturb structural probes
            2. Feed data through this module
            3. Evaluate via BERT
        ii. Setup perturbations via the syntax tree
            1. Setup Stanford Parser
            2. Setup module to perturb syntax tree (choose one easy rule to implement)
            3. Feed data through this module
            4. Evaluate via BERT
        iii. Design a detailed experiment that shows the influence of syntactic knowledge over the BERT accuracy
    b. Acquire BERT accuracy scores
    c. Analyze BERT accuracy scores