Brandon Chao
Tony Zhang
Daniel Cunningham

TTIC 31190 Project Proposal

Project/Task

We will replicate and extend the results of the paper “A Simple and Effective Approach to the Story Cloze Test” published by Srinivasan, Arora, and Riedl in March 2018 for our project. In this paper, the authors used a fully neural network approach trained on the last sentence of the prompt to achieve an accuracy within 1.1% of the state-of-the-art performance on this task. We will replicate these results using the described methodology in the paper, explore further methods to improve on this accuracy, and investigate why only using the final sentence instead of the full context provided higher accuracy.

Datasets/Evaluation

The Story Cloze test is a set of curated, crowdsourced 4-sentence stories, each paired with 2 potential ending sentences. One of these ending sentences is the labeled as the correct logical continuation of the story given the content and context. The stories themselves generally involve commonplace situations and are communicated via relatively straightforward sentence structures. The ROCStory dataset includes around 98,000 full stories (5-sentences) and 3,700 Story Cloze testing instances.

The best accuracy achieved on the test set by a model is 77.6% (Chaturvedi et al, 2017). Humans are able to achieve 100% accuracy.

Methods/Key Experiments

All the models are fit using feed-forward fully connected NNs with ReLU activations. The authors specify using 2 and 3 layers as well as the number of nodes per layer.

- “No Context”: Ignores the previous 4 sentences and trains only on the 5th. The idea is that there is something intrinsically unique about correct ending sentences.
- “Last Sentence”: Of the 4 context sentences, only the 4th is turned into an embedding and used to predict the correct ending. This is motivated by the idea that the ending sentence should flow most naturally from the sentence right before it.
- “Full Context”: Use all 4 context sentences, feed it into a GRU to get a 4800 dimensional embedding, and use that in a neural net.
- There appears to be a fair amount of freedom in the embeddings and the NN design itself.

Related Work

Original paper that established the data set: “A Corpus and Cloze Evaluation for Deeper Understanding of Commonsense Stories” (Mostafazadeh et al, 2016)

Current state-of-the-art performance: “Story Comprehension for Predicting What Happens Next” (Chaturvedi et al, 2017)

There are many additional papers about the Story Cloze Test, many of which are cited in the Srinivasan et al paper that we are planning to replicate. We will review the literature and implement ideas from some of these papers.

Timeline

1. Acquire the dataset - potentially clean and tokenize, etc.
    a. Probably use something out of the box
    b. Expected time: 3 days
2. Replicate the paper. Make sure we understand their methods and can achieve similar performance before trying to extend their methodology.
    a. Read/understand prerequisite knowledge and terminology
    b. Familiarize ourselves with the tools that the researchers used, install or replicate as necessary
    c. The paper is pretty clear about the NN structure. It should be pretty straightforward to set up all 3 of their core models in principle.
    d. We are unclear about computing power constraints. Some of our attempts may need to be reined in to run on laptops.
    e. Expected time: 4-6 days
3. Explore different options from improvement, borrow ideas from other papers
    a. Different NN types
    b. New embedding types
    c. Revisiting “simple” solutions like linear models
    d. Investigate why only using the final sentence in the context gives better accuracy
    e. Expected time: 3 days

The project workload will be split among the 3 group members evenly. We have not decided how to specifically split it up by task but we expect that each member will be involved in all aspects of the project.