Introduction to Deep Learning and Reinforcement Learning

Project Proposal

November 1st, 2019

# Group Members

Li, Yansong

Qu, Shuzheng

Su, Xuanyu

Yang, Siyuan

Linkletter, Maurice

# Statement of Problem

Our team has decided to compete in a superGLUE[1] task. The task we selected is **Choice Of Plausible Alternatives (COPA)** [2]. This task evaluates a model’s commonsense causal reasoning. It is tested with a series of questions where each question gives a premise and two plausible causes or effects. The correct choice is the alternative that is more plausible than the other.

**Approach**

# Siamese Manhattan LSTM: **Siamese**is the name of the general model architecture where the model consists of two identical subnetworks that compute some kind of representation vectors for two inputs and a distance measure is used to compute a score to estimate the similarity or difference of the inputs. In the attached figure1, the LSTMa and LSTMb share parameters (weights) and have identical structure. And Manhattan refers to the fact that they chose to use Manhattan distance to compare the final hidden states of two standard LSTM layers. Steps showed as following:

# Using tokenizer to extract word index from our words dictionary.

# Normalizing our data into word list with index number.

# Embedding our word index with pretrained glove word vector.

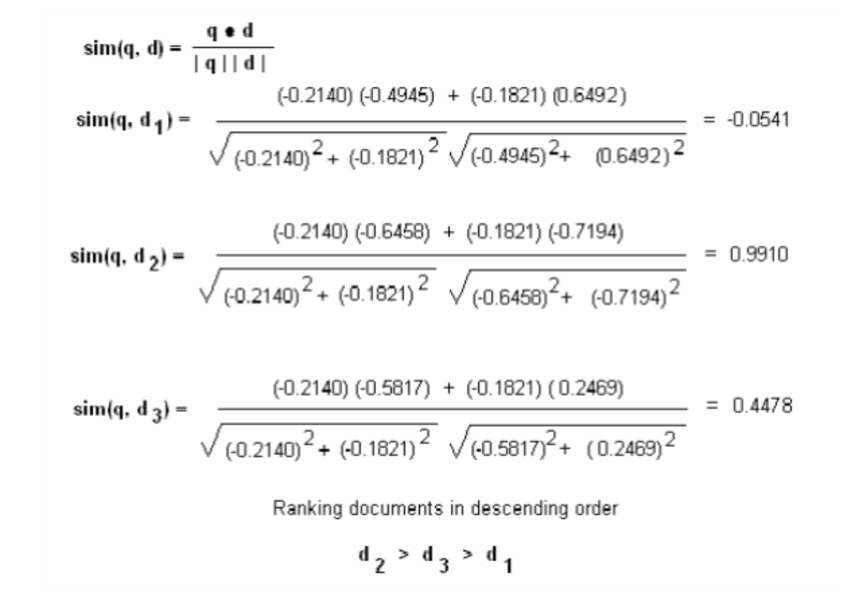
# Using two sub LSTM network separately train the sentences.

# Comparing the outputs of two LSTM as the similarity score.

# 

# *Figure1: Siamese Manhattan LSTM*

1. **Latent Semantic indexing**: Not directly comparing the cosine similarity of bag-of-word vectors, but first**reducing the dimensionality of our document vectors** by applying latent semantic analysis. LSA applies **principal component analysis** on our vector space and only keeps the **directions** in our vector space that contain **the most variance** (i.e. those directions in the space that change most rapidly, thus are assumed to contain more information). This is influenced by the **num\_topics** parameters we pass to the LsiModelconstructor. Steps showed as following:
   1. Set term weights(unique words dictionary) and construct the sentence matrix A and query matrix(question Q);
   2. Decompose matrix A and find the U, S and V matrixes, where A = U S (V)T (Implemented by scipy, svd);
   3. Implement a ‘*Rank 2 Approximation’* by keeping the first two columns of U and V and the first two columns and rows of S;
   4. Find the new document vector coordinates in this reduced 2-dimensional space;
   5. Find the **new query vector coordinates** in the **reduced 2-dimensional space**.



*Figure2 Similarity calculation*

# [Add a paragraph or two describing the proposed approach. For example, use using Manhattan distance between the premise and two plausible causes or effects. Or using BERT’s sentence classifier. Provide enough detail so that we can have 3 or 4 talking points when we meet with the professor. Please cite any papers or websites and add them to the “Reference” section.]

# Goals and Objectives

[What are our goals for this project? Maybe list 3 or 4 goals. For example:]

* Obtain the highest possible superGLUE score using state-or-the-art models and pre-trained transfer learning.
* Gain a deep understanding of Natural Language Processing problem space by compete in superglue tasks.

# Action Plan

**Sept 28th:** Project group formed.

**Oct 5th:**  First group meeting to review project choices.

**Oct 12th:** Second group meeting. Project Selection. Initial self-directed student topics identified.

**Oct 26th:** The team meet to discuss 3 possible approaches for the COPA superGLUE task.

**Oct 26th - Oct 30th:** Project proposal created with additional research on the three approaches.

**Nov 1st:** Review project proposal with professor.

**Nov 1st – Nov 15th:** Implemented our model.

**Nov 16th – Nov 17th:** Our first submitting to the superGLUE online benchmark.

**Nov 18th – Dec 1st:** Refine model and continue to submit to superGLUE.

**Nov 25th – Dec 2nd:** Create project poster.

**Dec 2nd – Dec 7th:** Create project report.

**Dec 5th:** Poster presentation

**Dec 8th:** Report due date

# References

[1] Alex Wang, Yada Pruksachatkun, Nikita Nangia, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. SuperGLUE: A stickier benchmark for general-purpose language understanding systems, 2019. arXiv:1905.00537.

[2] Roemmele, M., Bejan, C., and Gordon, A. (2011) Choice of Plausible Alternatives: An Evaluation of Commonsense Causal Reasoning. AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning, Stanford University, March 21-23, 2011.

[3] Adrien Sieg(2018, July 4). Test Similarity: Estimate the degree of similarity between two texts. Retrieved from <https://medium.com/@adriensieg/text-similarities-da019229c894>

[4] Eilor Cohen(2017, July 7). How to predict Quora Question pairs using Siamese Manhattan LSTM. Retrieved from <https://medium.com/mlreview/implementing-malstm-on-kaggles-quora-question-pairs-competition-8b31b0b16a07>