Project members

- Manasa Rangineni
- Revanth Reddy Bandaru
- Sowmya Davuluri

Project Title:

Question Answering on SQuAD 2.0

Objectives

- Building a model which can answer the questions provided with atmost accuracy
- Base line model is based on Bi-Directional Attention Flow (BiDAF)

- **Value:**
  - This model can stratify users who are looking to answer a specific question in Natural Language

Data

- Stanford Question Answering Dataset (SQuAD) is a reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or span, from the corresponding reading passage, or the question might be unanswerable.
- SQuAD2.0 combines the 100,000 questions in SQuAD1.1 with over 50,000 unanswerable questions written adversarially by crowdworkers to look similar to answerable ones. To do well on SQuAD2.0, systems must not only answer questions when possible, but also determine when no answer is supported by the paragraph and abstain from answering.

Approach

- converts the input word indices into word embedding for both the question and the context
- Uses a bi-Directional LSTM to cover the temporal dependencies between timesteps of the embedding layers outputs
- Modelling:
  - This uses a LTSM to take in the information between context representations given the background of the questions

Cont..

- Finally, we need to investigate how things will change due to the change in hyper parameters

Deliverables

- Python Code
- Final Report Pdf File
- Git Hub repository link
- Presentation(Youtube Video)

Evaluation methodology

- There are three metrics used for evaluating the model accuracy
- **Exact Match FM** : This metric measures the percentage of predictions that exactly match one of the ground truth answers.
- **F1 score** : This metric measures the average overlap between the predicted answer and the ground truth answers.
- **AvNA** : This stands for Answer vs No Answer and it measures the classification accuracy of the model when only considering its answer (any span predicted) vs no-answer predictions.

Citations

- https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1194/reports/default/15839661.pdf
- Seo, Minjoon, Kembhavi, Aniruddha, Farhadi, Ali and Hajishirzi, Hannanah. (2016) Bidirectional attention flow for machine comprehension. arXiv: 1611.01603, 2016.
- Rajpurkar, Pranav, Zhang, Jian, Lopyrev, Konstantin and Liang, Percy. (2016). Squad: 100,000+ quetions for machine comprehension of text. In Proceedings of the Conference of Empirical Methods in Natural Language Processing. CORR, abs/1606.05250, 2016.