Question Answering on SQuAD Dataset

NAMES OF THE MEMBER(S) OF THE GROUP

Disha Malik (B00842569); Nikitha Pasya(B00830710); Sneha Kotha(B00839933)

PROBLEM STATEMENT

In the given context, the model needs to find out the possible answers. A systematic approach for understanding the algorithms, techniques and systems around Question Answering is lacking so far. For this we are developing a model for documents in which information is factual, and the suitable answer is a single phrase.

POSSIBLE APPROACHES

Question Answering System uses science of information retrieval, natural language processing, and machine learning to answer the question automatically asked by user in natural language. A significant contributor to the advancement of Machine Comprehension models has been the availability of large datasets. The Stanford Question Answering Dataset(SQuAD),consists of questions posed by crowd workers on a set of Wikipedia articles, where the answer to each question is a segment of text from the corresponding reading passage. The data was collected in three stages: 1) Curating passages 2) Crowdsourcing QA on those passages and 3) Obtaining additional answers. Logistic regression model with a range of features, was implemented and its accuracy was compared with three baseline methods. Accuracy on this dataset was measured by F1 score and percent of exact matches through EM score[1].

The task of question answering has also gained a lot of interest in the computer vision community. Early works on visual question answering (VQA) involved encoding the question using an RNN, encoding the image using a CNN and combining them to answer the question [8][9]. Previous works in end-to-end machine comprehension use attention mechanisms. A dynamic attention mechanism, in which the attention weights are updated dynamically given the query and the context as well as the previous attention[5]. Simply using bilinear term for computing the attention weights in the same model drastically improves the accuracy[6]. Another mechanism includes reversing the direction of the attention (attending on query words as the context RNN progresses) for SQuAD[7].

Retrospective Reader distinguishes unanswerable questions in order to avoid giving plausible answers. Verification module, called verifier is used, in addition with encoder, to identify unanswerable questions. There are two stages to identify whether the question is answerable or not, 1) sketchy reading, briefly touches the relationship of passage and questions, and 2) Intensive Reading, verify answer and give final prediction. Model employs linear layer with SoftMax operations for final prediction. The implementation is based on BERT and ALBERT. The model uses the pre-trained LM weights in encoder module and available PLMs as encoder to build baseline MRC models: BERT and ALBERT. Two official metrics are used to evaluate the model performance: Exact Match (EM) and a softer metric F1 score, which measures the weighted average of the precision and recall rate at a character level. The model achieved 90% F1 score and 87% EM score on SQuAD dataset, outperforming the ALBERT baseline with simple threshold-based verifier, also achieves new state-of-the-art on the SQuAD[10].

Bidirectional attention flow mechanism is proposed to achieve state-of-the-art results on SQuAD dataset and CNN/DailyMail cloze test. BiDAF network is a multistage hierarchical process that represents the context at different levels of granularity. It includes character level, word level and contextual embeddings and uses query aware context representation. Use of attention mechanism offers some improvements: 1) use memory less attention mechanism which helps focus on learning attention between query and context and, 2) Use of attention mechanism in both directions, Q2C and C2Q providing complimentary information to each other. Model consists of six layers: 1) Character Embedding layer – uses character level CNNs. 2) Word Embedding Layer uses pre-trained embedding model GloVe. 3) Contextual Embedding Layer uses LSTM placed in both the directions. First three layers are applied to both query and context, and compute features from the query. 4) Attention Flow Layer links and fuses information from context and query words and reduce information loss. 5) Modeling Layer employs RNN to scan context and 6) Output Layer provides answer to query. The model achieved 73.3% EM score and 81% F1 score [9].

Project plan for the rest of the term

Milestone	Schedule
Preprocessing of data and implementation of word embedding layer	March 2
Implementation of RNN encoder and attention layer	March 10
Implementation of modeling and decoder layer	March 15
Improving the accuracy of the model using by hyperparameter tuning	March 20
Documentation	April 2

References

1. Pranav R., Jian Z., Konstantin L., and Percy L. Squad: 100,000+ questions for machine comprehension of text. EMNLP, 2016.
2. Matthew R., Christopher J., and Erin R. Mctest: A challenge dataset for the open-domain machine comprehension of text. EMNLP, 2013.
3. Karl Moritz H., Tomas K., Edward G., Lasse E., Will K., Mustafa S., and Phil B. Teaching machines to read and comprehend. NIPS, 2015.
4. Felix H., Antoine B., Sumit C., and Jason W. The goldilocks principle: Reading children's books with explicit memory representations. ICLR, 2016.
5. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. ICLR, 2015.
6. Danqi Chen, Jason Bolton, and Christopher D. Manning. A thorough examination of the cnn/daily mail reading comprehension task. ACL, 2016.Huohang Wang and Jing Jiang. Machine comprehension using match-lstm and answer pointer. arXiv preprint arXiv, 2016.
7. Mateusz M., Marcus R., and Mario F. Ask your neurons: A neural-based approach to answering questions about images. ICCV, 2015.
8. Stanislaw A., Aishwarya A., Jiasen L., Margaret M., Dhruv B., C.Lawrence Z., and Devi P. Vqa: Visual question answering. ICCV, 2015.
9. Minjoon S., Aniruddha K., Ali F., Hananneh H. Bi-directional attention flow for machine comprehension. arXiv, 2018.
10. Zhuosheng Z., Junjie Y.,Hai Z. Retrospective Reader for Machine Reading Comprehension. arXiv, 2020.