Analysis of Project

Nafis Tanveer Islam

Stanford Question Answering Dataset(SQuAD)**[1]** is a question answering dataset which consists of Wikipedia articles, Questions produced by crowd-sourcing on these wikipedia articles with answers to each question. SQuAD2.0 is the updated version of SQuAD1.1, where the original 100,000 questions are combined with over 50,000 unanswerable questions. The challenge/goal of this contest is to find the correct answer of a question from the given paragraph and also if the question is unanswerable, then the answer should be left blank. Here we selected the top three solutions that are available publicly and the following are a short overview for each of the procedures.

Zhang et al.**[2]** Proposed a retrospective reader method which includes a sketchy and intensive reading module. The sketchy reading module decides whether the question is answerable or not, and the intensive reading module predicts answers and answerability confidence. The input are paragraphs and questions that are represented as embedding vectors and then fed to a PrLM(Pre-Learned Language Model) encoder to get the embedded features. These embeddings are then fed to a multi-layer Transformer**[3]** for contextual representation where the last hidden layer, $\mathbf{h_n}$ where $\mathbf{H = \{h_1, h_1, h_1, \ldots h_n\}}$ is passed to a fully connected layer. This layer uses cross entropy loss to determine whether the question is answerable. Finally, the intensive reading Module produces the same hidden representation **H**. The **H** is then splitted into $\mathbf{H^P}$ $\mathbf{H^Q}$ for passage and questions which are fed into a Transformer based cross attention and a matching attention to produce the final output.

Lan et al.**[4]** used a method called ALBERT(A lite BERT) which is a pre trained language model with less parameters than the BERT**[5]** model by google. ALBERT achieves parameter downsizing by factorizing the embedding parameters into smaller matrices. So, instead of feeding the one-hot vectors into hidden layer H, initially it is projected into a low dimensional space of size E and then projects it to the hidden space. Thus the parameter is reduced to O(V * H) to O(V * E + E * H) where V is the size of vocabulary and H >> E. Their experimental results showed that, as the number of parameters are minimized, the accuracy is more than 88% with a significant drop in training time.

Zhang et al.**[6]** proposed a Syntex-Guided Network(SG-Net) Machine Reading Comprehension for question-answering of the **SQuAD** datasets. They used a transformer based Self-Attention Network(SAN) where they adopted a pre-trained dependent syntactic parse tree structure to produce the related node, integrating the relation as attentive guidance for enhancing SAN Transformer encoder. This SG Network is then applied to a pre-trained model BERT**[5]** which is then used for evaluating the challenge.

All of the above methods are used by multiple contestants achieving different rates of accuracy. All the methods proposed in these papers are different in their working procedure. The method proposed by Zhang**[2]** is the most similar to a human way of reading and comprehending a paragraph, which works by using a light reading method to determine whether the question is answerable or not, then using an intensive reading module to produce final answers. All the three models described above, use a pretrained network, where **[2]** and **[4]** used the pretrained model BERT by google and the last one**[6]** used pre-trained dependency parse tree structure. Zhang**[6]** also used BERT to compare their results against their own model and showed that their model achieves substantial performance updates against the BERT model. Also from the challenge leaderboard we can see that the top 20+ models perform significantly better than the Human Performance on Exact Match and F1 score.

References:

[1] Rajpurkar, P.; Jia, R.; and Liang, P. 2018. Know What You Don’t Know: Unanswerable Questions for SQuAD. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 784–789.

[2] Zhang, Zhuosheng, Junjie Yang, and Hai Zhao. "Retrospective reader for machine reading comprehension." *arXiv preprint arXiv:2001.09694* (2020).

[3] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention is All you Need. Advances in neural information processing systems 30: 5998–6008.
[4] Lan, Zhenzhong, et al. "Albert: A lite bert for self-supervised learning of language representations." *arXiv preprint arXiv:1909.11942* (2019).
[5] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).
[6] Zhang, Zhuosheng, et al. "Sg-net: Syntax-guided machine reading comprehension." *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 05. 2020.

Paper and Code Link:

Paper: https://arxiv.org/abs/2001.09694
Code: https://github.com/cooelf/ AwesomeMRC

Paper: https://arxiv.org/abs/1909.11942
Code: google-research/albert: ALBERT: A Lite BERT for Self-supervised Learning of Language Representations

Paper: https://arxiv.org/abs/1908.05147
Code: https://github.com/cooelf/SG-Net