

Paper Title: Educational Multi-Question Generation for Reading Comprehension

Paper Link: <https://aclanthology.org/2022.bea-1.26.pdf>

1. Summary

1.1 Motivation/ Purpose/ Aim

The authors propose a novel approach to generating multiple question wordings for a given passage in order to enhance reading comprehension instruction. Their method utilizes a recurrent neural network model that is trained on a dataset of questions and corresponding passages. The model learns to generate diverse and informative questions that assess understanding of the text at multiple levels. The authors demonstrate the effectiveness of their approach through an empirical evaluation on a benchmark question generation dataset. Their results show that their model can generate questions that are as good as or better than those generated by human experts. The authors also discuss the potential benefits of using their approach in educational applications.

1.2 Contribution

1. They propose a new task, multi-question generation, for assessing reading comprehension. Multi-question generation involves generating multiple questions that assess the same concept, but with different wording.
2. They develop an evaluation framework for multi-question generation. This framework ensures that the generated questions are both semantically similar and lexically diverse. This is important because it means that the questions assess the same concept, but they do not simply repeat the same words and phrases from the passage.
3. They compare multiple question generation approaches in a two-question generation condition. This experiment shows that there is a trade-off between question answerability and lexical diversity between the two questions. This is an important finding because it suggests that there is no single best way to generate multiple questions for a given passage
4. The authors conduct an empirical evaluation of MQG models on a benchmark question generation dataset. Their results show that their baseline MQG model can generate questions that are as good as or better than those generated by human experts.

1.3 Methodology

Dataset collection and labeling were performed initially. The authors then extracted features from the paragraph and represented them in a format that the neural network model could understand using a variety of techniques. Among these methods were named entity recognition, part-of-speech labeling, and word embedding. In addition, the authors devised a ProphetNet and pre-trained SBERT model with the purpose of multi-Question generation. To measure the semantic similarity between generated questions to ensure that the questions assess the same content, they used SBERT model, while the ProphetNet is utilized to transform a single question

output into a paraphrased second question, fine-tuning to output two questions sequentially. In conclusion, the researchers assessed the effectiveness of the suggested model by applying the subsequent metrics to the gathered data set: F1 score. In addition, the authors conducted experiments utilizing various architectures of neural networks. On the collected dataset, they discovered that the proposed SBERT model outperformed the other architectures.

1.4 Conclusion

The authors end the research by claiming that the suggested recurrent neural network model is a viable solution for the question - answering from a text sample. The model performed well on the obtained data, suggesting it might be utilized to aid in the creation of a question - answering for the educational institutions.

2 Limitations

2.1 First Limitation

The evaluation framework developed by the authors focuses heavily on question answerability, which measures whether a question can be correctly answered based on the given passage. However, this metric does not capture other important aspects of question quality, such as the question's ability to promote deeper understanding, critical thinking, and creativity.

2.2 Second Limitation

The authors discuss the potential benefits of using MQG in educational applications, but they do not explore these applications in depth. More research is needed to determine how MQG can be effectively integrated into different teaching and learning contexts.

3 Synthesis/ Future work

1. Evaluating the proposed model on other datasets to assess its generalizability.
2. Enhancing the system by including teacher evaluation to collect education-specific feedback on sets of questions and our desirable question properties.
3. Implementing desirable question metrics in a reinforcement learning objective to produce higher quality questions, similar to previous work in abstractive summarization and text simplification
4. Implementing more advanced paraphrase systems, such as the syntax-aware system proposed, could be leveraged for this work.