Question Generation

Prajneya Kumar: 2019114011
Shivansh S: 2019114003

Advanced NLP: Project Write Up

1 Introduction

In the age of technology, it is natural to ingrain it with all possible fields, including education. With the ever growing amount of educational content it is becoming more and more difficult to generate multiple questions manually. Hence, our project works on question generation based on educational content which will automatically generate questions based on text data as input.

2 Literature Review

Majority of the research work carried out so far has been focused on three models:

1. Rule Based Generation Model
2. LSTM + Attention + Linguistic Features Model
3. The QG-Net Model

Apart from the rule-based model, we explore each one of the two NN-based models, as explained below:

2.1 Learning to Ask: Neural Question Generation for Reading Comprehension

This paper focuses on generating questions for sentences from text passages in reading comprehension by introducing a LSTM+Attention+Linguistic Features model. Seq2Seq Learning is used to train the model and is shown to significantly outperform earlier rule-based models. [1]

2.2 QG-Net: A Data-Driven Question Generation Model for Educational Content

This paper introduces a RNN-based model, particularly built for the generation of quiz questions from educational content. The paper compares the results of the earlier model (LSTM+Attention+Linguistic Features) with QG-Net and proves that it outperforms the same. [2]

3 Our Methodology

We aim to implement the following models:

1. Baseline: LSTM + Attention + Linguistic Features Model
2. Baseline+: The QG-Net Model (trained on GLoVe)
3. Baseline++: The QG-Net Model (trained on BERT)

First part of our problem deals with basic Question Generation, and how well they can be generalised. We will explore how well you can generalise LSTM based encoder-decoder model trained on SQuAD dataset onto question generation based on Educational Data.

Second part of our project will be focussed on improving upon our baseline, both based on better word representations and better generalibility. We would use contextual word embeddings like BERT instead of basic ones like GloVE for the former, and for the latter we would try implementing QGNet, a SoTA generalizable model for question generation

References

[1] Xinya Du, Junru Shao, and Claire Cardie. *Learning to Ask: Neural Question Generation for Reading Comprehension*. 2017. arXiv: 1705.00106 [cs.CL].

[2] Zichao Wang et al. "QG-net: a data-driven question generation model for educational content". In: June 2018, pp. 1–10. DOI: 10.1145/3231644.3231654.

*****