# Automatic Question Generation

Sasi Kiran Gaddipati, Desiana Dien Nurchalifah

June 1, 2020

## 1 Topic

The evaluator takes considerable time to prepare questions and evaluate answers, when examining to assess students. We aim to automatically generate questions through its associated answers [1]. The application is also considered secure, compared to the manual evaluator making questions, which may lead to exposure to an unauthorized party. We would like to train on the SQuAD dataset, the answer as input and the corresponding question as output.

## 2 Techniques in NLP

The techniques that we consider using are as follows:

- Tokenization: To tokenize the words in questions and answers
- Named Entity Recognition: To extract the words information from sentences
- Seq2Seq Learning: To train the input and output words
- Positional Encoding: To encode the position of the answer

## 3 Evaluation Strategy

According to the work of [1], evaluation is done in two methods:

- Human evaluation: Judges measure quality of generated questions according to evaluation criteria, semantics, context and grammar. To check the agreement among judges, Fleiss' Kappa score [2] is calculated.
- Automatic evaluation: BLEU-4 scores are used to evaluate rare words problem. In addition of BLEU, METEOR[3] is proposed to compare top rated questions performance.

## References

[1] Qingyu Zhou, Nan Yang, Furu Wei, Chuanqi Tan, Hangbo Bao, and Ming Zhou. Neural question generation from text: A preliminary study. *CoRR*, abs/1704.01792, 2017.

[2] JL Fleiss. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378—382, November 1971.

[3] Satanjeev Banerjee and Alon Lavie. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.