PM Questions NLP
Jacob Löbkens
Alejandra Valdez

Abstract

In our project, we aim to show which type of influence a change in the question has on the output of a question answering model.
In order to achieve this, we will generate adversarial examples and simulate speed-reading the question by changing the input.

[1] generates adversarial examples for a range of NLP-tasks but not for question answering.
Therefore, we want to generate adversarial examples for a question answering model. Similar to the mentioned paper we will change verbs, nouns, and adjectives in a question to their most immediate synonyms. Furthermore, we will contrast these results.

We will also try to incorporate the findings of [2] in our project.
The paper implemented a "speed-reading" algorithm in the model. The model is able to decide how much to read from the input to solve a task. Because the architecture to achieve this is complicated, we will simulate the speed-reading by skipping a set amount of words in the questions.

For each of of the changed datasets we will train a separate model. Additionally, we will train a baseline model on a unchanged dataset.
The models will be a very basic architecture. We use a GRU like [3] implement for the bAbI dataset in Keras.

The trained models will be evaluated and compared using a given evaluation-script.

As dataset we will use SquAD [4]. This dataset offers its own evaluation-script so that we are able to compare our results in a leaderboard.

References:

[1] http://aclweb.org/anthology/D18-1316 // Generating Natural Language Adversarial Examples
[2] https://tsujuifu.github.io/pubs/emnlp18_lstm-shuttle.pdf // Speed Reading: Learning to Read ForBackward via Shuttle
[3] https://cs224d.stanford.edu/reports/StrohMathur.pdf // Question Answering using Deep Learning
[4] https://rajpurkar.github.io/SQuAD-explorer/