

CS4248 Natural Language Processing

AY 2023/24 Semester 1

Pre-defined project topic: Machine reading comprehension

This task involves implementing a question answering system such that given a passage and a question in English, the system provides an answer to the question based on the passage, where the answer is a segment of text (called a span) from the passage.

We will use SQuAD (Stanford Question Answering Dataset) in this project for training and testing. The training data (train-v1.1.json) and test data (dev-v1.1.json) have been released on Canvas. Details of this task and the dataset can be found in the following paper:

SQuAD: 100,000+ Questions for Machine Comprehension of Text
Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, Percy Liang
Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing
<https://aclanthology.org/D16-1264.pdf>

Evaluation of your implemented QA system is to be carried out using the official evaluation script evaluate-v2.0.py:

<https://worksheets.codalab.org/rest/bundles/0x6b567e1cf2e041ec80d7098f031c5c9e/contents/blob/>

Two evaluation metrics are used:

Exact match: This metric measures the percentage of your system's answers that *exactly* match any one of the gold-standard answers.

F1 score (macro-averaged): This metric measures the maximum F1 score based on the *overlap* between the system's answer and a gold-standard answer (treated as bags of tokens) for a question, and then averages the F1 scores over all questions.

Other final project topics

Other project topics are also welcome (subject to approval by the lecturer). Some sample final project topics in the past years include:

Aspect-based sentiment analysis
Grammatical error correction
Opinion summarization