Scalable Machine Learning and Deep Learning - Project Proposal
Deep learning-based question answering system
Deadline: November 29, 2020
Jaweriah_Xing

Jaweriah Alvi <jaweriah@kth.se>
Xing Zeng <xingzeng@kth.se>

Project Description:
Lots of information is now available on the web. We mainly search on the web to find answers to our questions. Our main idea is to create a Question Answer System.

Tools:
TensorFlow, Google Colab.

Data:
We will be using **S**tanford **Qu**estion **A**nswering **D**ataset (SQUAD). It is a reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or span, from the corresponding reading passage, or the question might be unanswerable [3].

Methodology and algorithm:
Our implementation will be based on the deep learning approach and will use Recurrent Neural Networks(RNN) with Bidirectional Attention Flow (BiDAF).
We will first convert the QA Dataset to word embedding in Embedding Layer. Our next step would be to encode the dataset, here we will use an RNN (GRU/LSTM) based encoder. Then we will use the attention layer to find the match between the hidden vector for the dataset and the hidden vector for a question, then we will use Bidirectional Attention Flow (BiDAF) to compute the Similarity Matrix. Finally, we will use Context to Question Attention (C2Q), Question to Context Attention (Q2C), and combine them. The Output Layer will show the predicted answer.

References:
[1] https://www.sciencedirect.com/science/article/pii/S1877050918308226
[2] https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00341-6
[3] https://rajpurkar.github.io/SQuAD-explorer/