Project #5 for the Biomedical Information Retrieval Course

Due: Dec 12, 2023

General Guideline

This homework is basically an individual homework. Each student has to do it all by himself (or herself). The final score will be evaluated from the system performance and individual demonstration.

Homework Overview

Implement training of the Bidirectional Encoder Representations from Transformers Technique (BERT) for a set of Drug-Drug Interaction sentences from SemEval-2013 Task 9. In this project, you have to train the computational BERT model(s) to determine the category of the test DDI-pair, as an NLP classification task. The training text set could be chosen from xml files in the Train sub-directory of the GitHub site in System Description, whereas the test set in the Test-DDI sub-directory. Only 2-entity (i.e. pair DDI) needs to be used. You have to preprocess the text set from Train collection, which describe the types of DDI, specifically as “effect / mechanism / int / advise / none (i.e. False, without DDI). Then, train the BERT language model to implement the NLP classification task of PPI-pair using the training set of expert-annotated DDI-pair sentences.

System Description

1. I suggest that you can obtain the train set from the following site https://github.com/albertrial/SemEval-2013-task-9/tree/master/data
2. In the final evaluation, each individual should present BERT training, running results, and system demo to verify classification task performance.