Alžbeta Strompová 492945

Second Term Project: CQADupStack Collection: Project Report

System Description

My IR system uses the sentence transformer from https://www.sbert.net/ , which is a pre-trained model that has been fine-tuned for many use cases. It was trained on a large and diverse dataset of over 1 billion training pairs.

Training & System Parameters

For sentence transformer, I used *all-mpnet-base-v2.* After that I trained model on prepare *train_judgements*. I use *MultipleNegativesRankingLoss*. It is a loss function if you only have positive pairs, for example pairs of (query, response). I choose number of epoch to 1 and warmup steps to 100. Warmup step represents the period during which the learning rate is increased from 0 to the maximal learning rate, after which the learning rate is linearly decreased back to zero. (more: https://www.sbert.net/docs/training/overview.html)

Development Process & Other Interesting Findings

In addition to reweighting the titles, I've also experimented with changing number of epoch and warmup step.

I have also tried adding BM25Plus as re-ranker system, but I found this to lower the resulting percentage.