

Approach 1:

We encode the papers with sentence transformers and use cosine similarities to find the most similar sample for every paper.

Approach 2:

We train a neural network to take in as input the two concatenated embeddings of a pair of sentences and give them a probability of being good for each other. During training, all pairs of daily papers are given a score of 0, and all pairs of papers and positive samples are given a score of 1. Then, for every daily paper, we select the positive sample that maximizes the score with the paper. We use a simple FFN and train it for 50 epochs using an Adam optimizer with an lr of 0.001.