Prof. Dr. Goran Glavaš,
M.Sc. Fabian David Schmidt
M.Sc. Benedikt Ebing
Lecture Chair XII for Natural Language Processing, Universität Würzburg

8. Exercise for "Multilingual Natural Language Processing"

19.07.2024

1 Paper Readings

We will focus on four papers on sentence-level representation learning.

- Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks
- SimCSE: Simple Contrastive Learning of Sentence Embeddings
- Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation
- Language-agnostic BERT Sentence Embedding

2 Supervised Representation Learning

1. Explain the training objective of the original Sentence-BERT transformer. Why does the objective enable cosine similarity search at inference time?

2. Can you think of intuitions as to *why* SRoBerta *does not* outperform SBERT, in contrast to other types of downstream tasks?

3 Self-Supervised Representation Learning

1. Briefly explain the core idea of contrastive learning and how the training objective is typically constructed.

2. How does unsupervised SimCSE learn sentence-level representations in a self-supervised fashion? How does it thereby improve over other potentially self-supervised objectives?

3. Imagine you want to train your own multilingual sentence transformer. List and briefly explain some key considerations in scaling up the training procedure.

4 Knowledge Distillation For Representation Learning

1. What is knowledge distillation and how does it work (on the case of multilingual sentence transformers)?

2. Can you think of and elaborate on factors that affect how well knowledge distillation works?

5 Additional Exercises (Not required for bonus)

1. Your are given the following two embedding pairs from a bi-encoder. Compute the InfoNCE loss with cosine similarity and temperature T=0.5 as shown in the lecture slides.

$$\text{Positive pair} \rightarrow [0.8109, -0.9391, 0.2519], [-1.2887, 1.5057, 0.4449]$$
$$\text{Negative pair} \rightarrow [0.8109, -0.9391, 0.2519], [2.1968, 0.4785, 1.5207]$$