Marián Ligocký

Answer Correctness Estimation on a Question

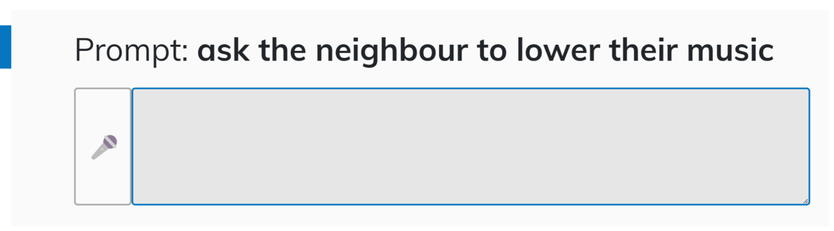
BRNO UNIVERSITY OF TECHNOLOGY — FACULTY OF INFORMATION TECHNOLOGY

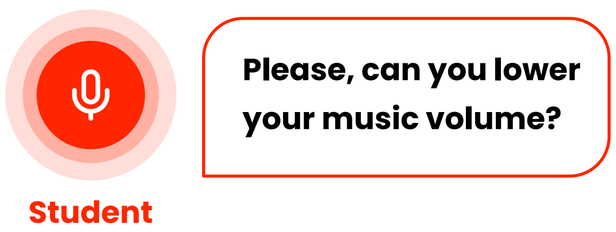
Excel@FIT 2023

Supervisor: **Ing. Igor Szőke, Ph.D.**

01 Problem

English learning app:





PREDEFINED CORRECT ANSWEARS

- Could you please lower your music volume?
- Would you lower the volume of your music please?
- Would you please turn down the volume of your music?

EVALUATION

by a literal comparison => **incorrect**,
because it is not among correct answers

by a NLP processing => **correct**
Model **finds similarity** to one/many of correct answers

02 Embeddings = vectors

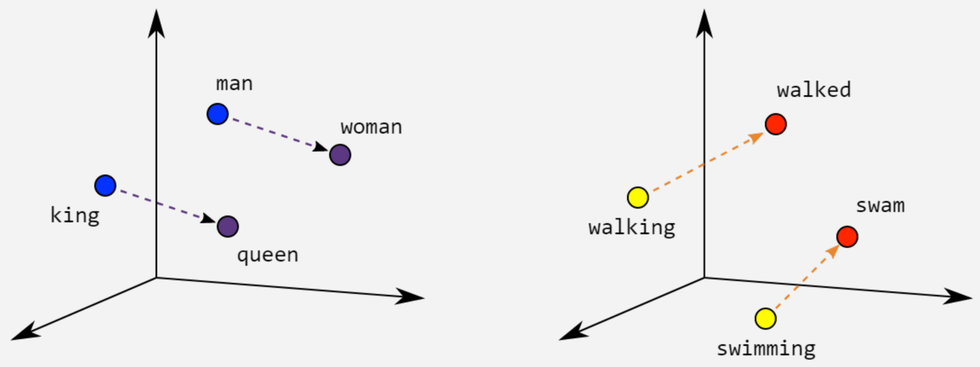


Figure 1, embedding encodes meaning of words and relationships

03 NLP tasks

STS - SEMANTIC TEXTUAL SIMILARITY

Are topics similar?

Can you lower your music?	**Bi-encoder**	**Cross-encoder**
Can you increase your music?	0.753	0.462
Can she increase your music?	0.602	0.378
Please might you turn volume down?	0.503	0.602
Would you lower your music?	0.873	0.892

Table 1, bi-encoder: all-MiniLM-L6-v2, cross-encoder: roberta-large

NLI - NATURAL LANGUAGE INFERENCE

Given first sentence, is second sentence entailnment, contradiction or neutral?

Can you lower your music?	**NLI**
Can you increase your music?	❌ contradiction
Can she increase your music?	❌ contradiction
Please might you turn volume down?	✅ entailnment
Would you lower your music?	✅ entailnment

Table 2, model: nli-deberta-v3-base

04 Model architecture

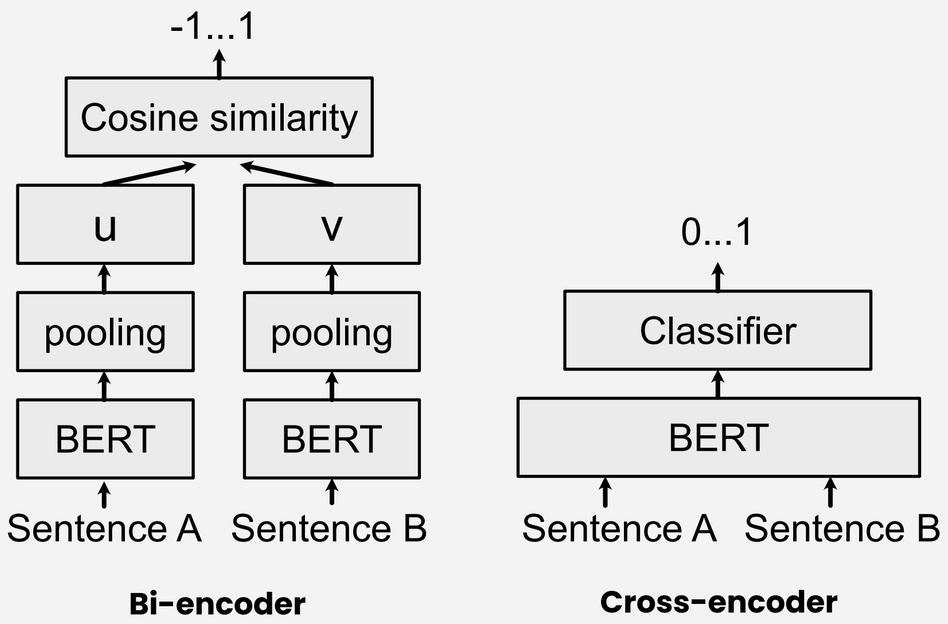


Figure 2, comparison of two approaches to get similarity (SBERT).

05 Future

- more human-annotated data
- grammatical model
- large language models

05 Results

- +13.6% Pearson correlation
- two dimensions: **label** + **similarity score**

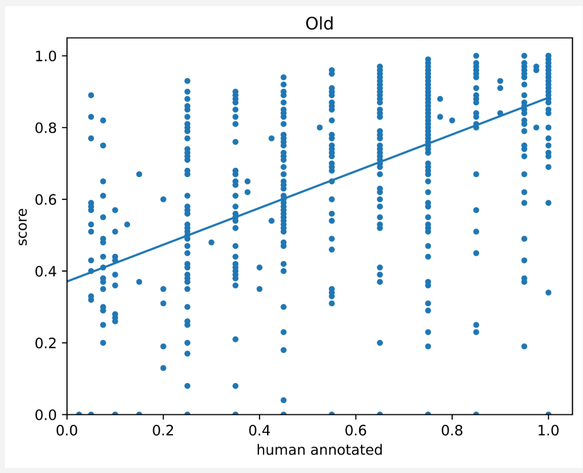


Figure 3, old scoring system

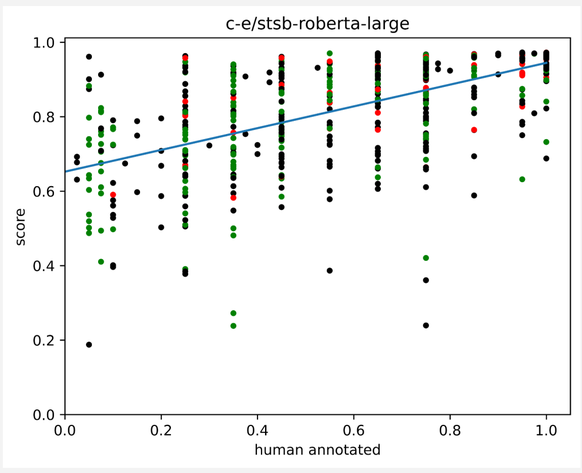


Figure 4, our model with NLI labels