Enhancing Expressive TTS Synthesis for Multilingual Low-Resource Languages: Challenges and Applications

Daniel Hládek, Michal Stromko, Kristián Sopkovič, Matúš Pleva, Ming, Hsiang Su

Department of Electronics and Multimedia Communications TUKE

daniel.hladek@tuke.sk

Abstract

We introduce a novel benchmark for information retrieval in the Slovak language, with a unique twist that could be of particular interest to researchers in speech recognition. By leveraging a question-answering dataset, we fine-tune and evaluate sentence transformers with direct implications for speech-based applications. The dataset, named Retrieval SkQuAD, has been integrated into two prominent evaluation frameworks, MTEB and BEIR. It consists of 19,000 manually annotated answers to 1,134 questions, each rated with a relevance score, alongside metadata on document utility for answer generation. What sets this apart from typical question-answering datasets is its nuanced assessment of partial relevance, critical for handling responses derived from multiple spoken documents.

We specifically fine-tuned sentence transformers and BERT-based models for retrieving documents that contain the correct answers within Slovak Wikipedia, a valuable resource for speech-to-text systems. Our fine-tuning process leverages adversarial questions as hard negatives, mimicking the challenges of spoken dialogue with ambiguous queries, and delivering substantial gains in retrieval accuracy. The results demonstrate that our approach significantly improves state-of-the-art performance in Slovak information retrieval, offering potential applications for more accurate spoken language systems and enhancing the robustness of speech-based question-answering models.