OpenFact at CheckThat! 2024: Cross-Lingual Transfer Learning for Check-Worthiness Detection

POZNAŃ UNIVERSITY OF ECONOMICS AND BUSINESS

Marcin Sawiński, Krzysztof Węcel, Ewelina Księżniak

Poznań University of Economics and Business, Department of Information Systems
Al. Niepodległości 10, 61-875 Poznań, Poland

Abstract

Several mono- and multilingual pre-trained language models were fine-tuned using different variants of the training datasets. Cross-lingual transfer learning was applied without instance transfer and proved to be effective for Arabic and Dutch. Additionally, we tested the effectiveness of class balancing using several under-sampling methods, which, when combined with appropriate model selection and cross-lingual transfer learning, produced the second-best results for Arabic and English.

Three parts of the study:
1. Finding the best monolingual model to use as a baseline.
2. Preparing multilingual training dataset variants.
3. Training and evaluating mono- and multilingual models on the prepared datasets.

Main driving research questions:
RQ3: How can cross-lingual transfer be leveraged to improve check-worthiness detection using training data in multiple languages?
RQ4: Is it possible to outperform random under-sampling with methods informed by annotation quality or training dynamics?

Models used:
English: DeBERTa V3 base, DeBERTa V3 large
Arabic: CAMeLBERT MSA, CAMeLBERT DA, CAMeLBERT CA
Dutch: RobBERT 2023 large, BERTje
Multilingual: mDeBERTa V3 base, XLM-RoBERTa base

Phases of the experiments:
1. Testing single language models using unaltered datasets.
2. Testing cross-lingual transfer learning using various concatenations of datasets.
3. Testing the impact of various structural changes to the training datasets.

Under-Sampling methods:
- **RUS** – random under-sampling.
- **DUS** – Symmetrically removing the most easy-to-learn and hard-to-learn examples. All majority class examples were sorted in descending order by their ℓ2 distance from the reference point (variability, confidence)=(0.5, 0.5) and removed until the desired class count was reached.
- **HUS** – First removing all hard-to-learn examples (defined as examples having an ℓ2 distance from (variability, confidence)=(0.5, 0.5) greater than 0.35 while having a confidence < 0.5), and then removing easy-to-learn examples sorted by descending distance from (variability, confidence)=(0.5, 0.5) until the desired class count was reached.
- **CUS** – First removing all examples from the majority class with correctness less than five, and later, if necessary, randomly choosing examples with correctness equal to five until the desired class count was reached.

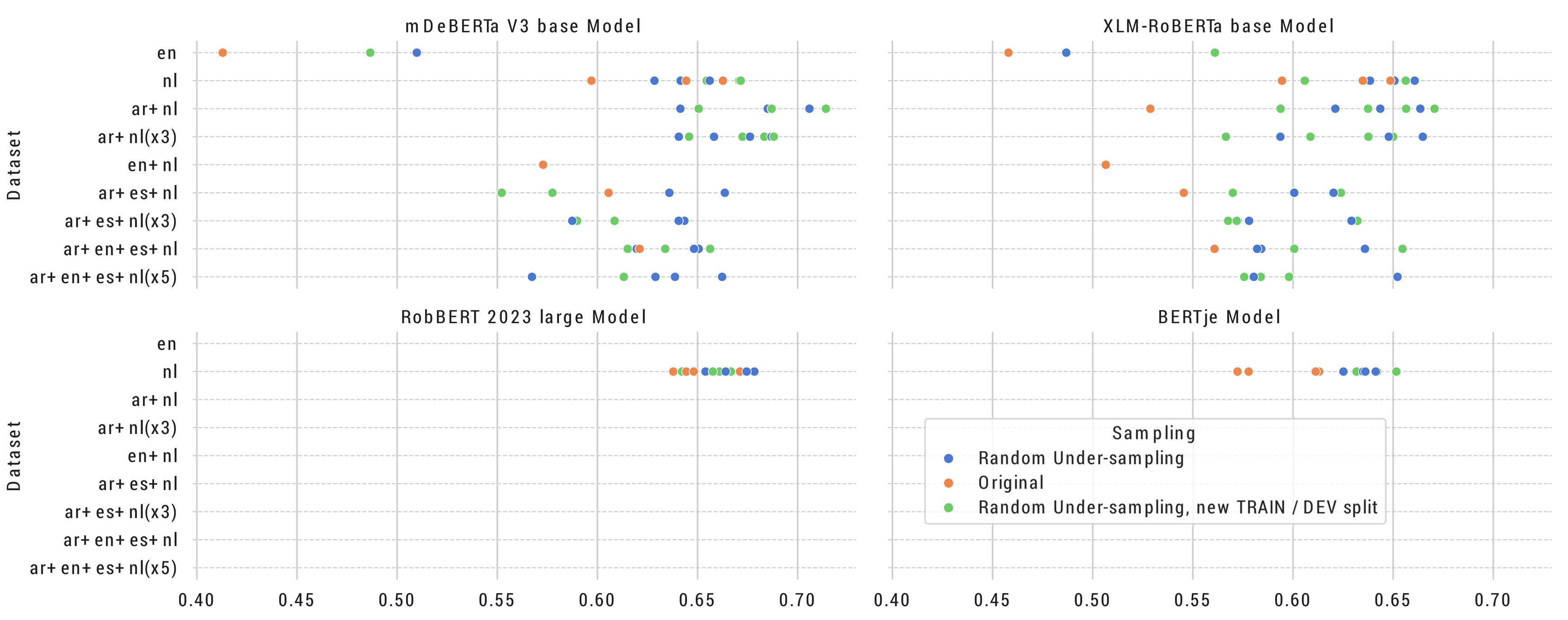

Figure: Results of Cross-Lingual Transfer experiments - F1 score (positive class) for Dutch dev_test dataset.

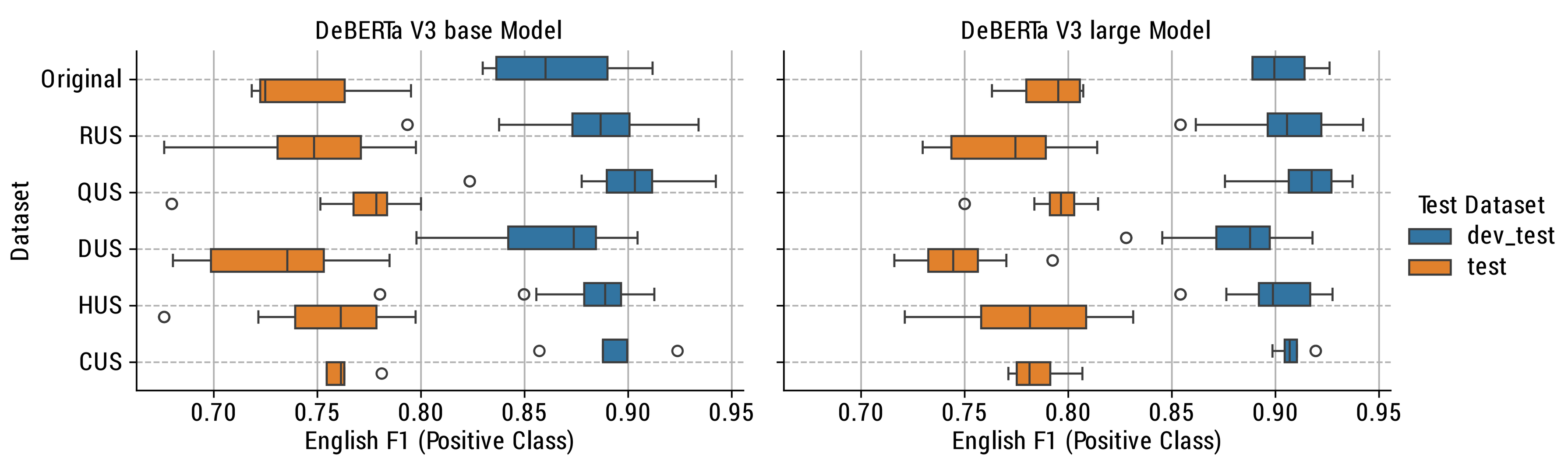

Figure: Results of under-sampling experiments - F1 score (Positive Class) for English dev_test dataset.

NCBR National Centre for Research and Development

Program strategiczny INFOSTRATEG

This research is supported by the grant „OpenFact - artificial intelligence tools for verification of the veracity of information sources and fake news detection" (INFOSTRATEG-I/0035/2021-00), within the INFOSTRATEG I program of the National Center for Research and Development, under the topic: Verifying information sources and detecting fake news.

Minister of Science Republic of Poland

Regionalna Inicjatywa Doskonałości

Supported by funds granted by the Minister of Science of the Republic of Poland under the „Regional Initiative for Excellence" Programme for the implementation of the project "Poznań University of Economics and Business for Economy 5.0: Regional Initiative – Global Effects (IREG)".