How Deep Contextualized Representation and Attention Mechanism Justifies Explainable Cross-lingual Sentiment Analysis

Rouzbeh Ghasemi, Saeedeh Momtazi*

Computer Engineering Department, Amirkabir University of Technology, Tehran, Iran

Abstract

The number of applications in sentiment analysis is growing daily, and research in this field is increasing. By analyzing user sentiment, companies gain valuable information about needs, thoughts, reactions, and ideas to finding the strengths and weaknesses of products or services. Despite the rapid growth of data sources in English, low-resource languages suffer from the lack of data for accurate training models. Moreover, providing no explanation about the output, users cannot trust such systems. In this study, we propose a cross-lingual deep neural model to improve the accuracy of sentiment analysis for low-resource languages. For training cross-lingual models, we used English Amazon dataset and Persian Digikala dataset while using pre-trained cross-lingual language models, XLM-RoBERTa, a contextualized transformer-based word representation. In addition, to improve the explainability of the model as well as its performance, we propose a cross-lingual attention model with Long Short Term Memory (LSTM) network to detect the informative words which have an impact on the positive and negative polarity of the user's review. With two different classifications, we show the superiority of the proposed model compared to the state-of-the-art monolingual techniques and cross-lingual models. The results show 0.55% improvement compared to the cross-lingual sentiment analysis proposed by Ghasemi et al. (2020) and 15.08% improvement compared to

*Corresponding author
Email address: momtazi@aut.ac.ir (Saeedeh Momtazi)