Document-level Sentiment Analysis: An Exploratory Case Study of Low-resource Language Urdu

Authors : Ammarah Irum, Muhammad Ali Tahir

Abstract : Document-level sentiment analysis in Urdu is a challenging Natural Language Processing (NLP) task due to the difficulty of working with lengthy texts in a language with constrained resources. Deep learning models, which are complex neural network architectures, are well-suited to text-based applications in addition to data formats like audio, image, and video. To investigate the potential of deep learning for Urdu sentiment analysis, we implemented five different deep learning models, including Bidirectional Long Short Term Memory (BiLSTM), Convolutional Neural Network (CNN), Convolutional Neural Network with Bidirectional Long Short Term Memory (CNN-BiLSTM), and Bidirectional Encoder Representation from Transformer (BERT). In this study, we developed a hybrid deep learning model called BiLSTM-Single Layer Multi Filter Convolutional Neural Network (BiLSTM-SLMFCNN) by fusing BiLSTM and CNN architecture. The proposed and baseline techniques are applied on Urdu Customer Support data set and IMDB Urdu movie review data set by using pre-trained Urdu word embedding that are suitable for sentiment analysis at the document level. Results of these techniques are evaluated and our proposed model outperforms all other deep learning techniques for Urdu sentiment analysis. BiLSTM-SLMFCNN outperformed the baseline deep learning models and achieved 83%, 79%, 83% and 94% accuracy on small, medium and large sized IMDB Urdu movie review data set and Urdu Customer Support data set respectively.

Keywords : urdu sentiment analysis, deep learning, natural language processing, opinion mining, low-resource language
Conference Title : ICNLPII 2023 : International Conference on Natural Language Processing and Intelligent Interfaces
Conference Location : Karachi, Pakistan
Conference Dates : December 25-26, 2023