

Linguistic Features for Sentence Difficulty Prediction in Aspect-Based Sentiment Analysis

Authors : Adrian-Gabriel Chifu, Sebastien Fournier

Abstract : One of the challenges of natural language understanding is to deal with the subjectivity of sentences, which may express opinions and emotions that add layers of complexity and nuance. Sentiment analysis is a field that aims to extract and analyze these subjective elements from text, and it can be applied at different levels of granularity, such as document, paragraph, sentence, or aspect. Aspect-based sentiment analysis is a well-studied topic with many available data sets and models. However, there is no clear definition of what makes a sentence difficult for aspect-based sentiment analysis. In this paper, we explore this question by conducting an experiment with three data sets: "Laptops", "Restaurants", and "MTSC" (Multi-Target-dependent Sentiment Classification), and a merged version of these three datasets. We study the impact of domain diversity and syntactic diversity on difficulty. We use a combination of classifiers to identify the most difficult sentences and analyze their characteristics. We employ two ways of defining sentence difficulty. The first one is binary and labels a sentence as difficult if the classifiers fail to correctly predict the sentiment polarity. The second one is a six-level scale based on how many of the top five best-performing classifiers can correctly predict the sentiment polarity. We also define 9 linguistic features that, combined, aim at estimating the difficulty at sentence level.

Keywords : sentiment analysis, difficulty, classification, machine learning

Conference Title : ICIRA 2024 : International Conference on Information Retrieval and Applications

Conference Location : Tokyo, Japan

Conference Dates : July 22-23, 2024