Abstract

With the large-scale penetration of social media into our daily lives, it has become a platform for individuals to share and express their views, feelings, opinions, and thoughts. Identifying emotions has many applications ranging from personalized marketing to behavior study. Individuals express their feelings in a language that is frequently accompanied by ambiguity and figure of speech, which makes it difficult even for humans to comprehend. In this paper, we propose a new approach to classify text into emotion categories. We use Twitter data as labeled input, this data is labeled using hashtags and addresses features like emoticons, emoji, apostrophes, Twitter slang and spelling variations which are a part of informal language on social media. Our model uses word vectors generated by architecture like Word2vec, Glove, and Fasttext to generate word representations of the text. We then investigate the utility of these models on random forest classifier. Ultimately we compare the results to find the best model for text classification based on emotions. We achieve an overall 91% precision for four emotional classes on a mined dataset of more than 100,000 tweets. This is a very useful tool to understand human behavior and a natural step beyond the positive/negative polarity.
References


Index Terms

Computer Science Information Sciences
Keywords

Keywords Word vectors, random forests, Word2vec, Glove, emotion, text classification