Abstract

Sentiment analysis is an application of natural language processing. It is also known as emotion extraction or opinion mining. There is a vast growth in use of social networking sites and microblogging websites, by which communication between people from different countries, cultures and psychological and physical backgrounds became more direct, resulting in more and more conflicts about thoughts and speech used between these people. Hate speech can be explained and it can be put into context as the use of aggressive, violent or offensive language, which targets a specific group of people who shares a common property. In India, a lot of hate speech generated posts are placed on social networking sites. So, to block and catch the hate speech generated posts and to avoid necessary conflicts, we thought of introducing a technique using Twitter, as our data source. We will create and analyze datasets that can be used for machine learning. We are going to analyze the negative and positive hate speech tweets which are generated, and use machine learning algorithms to analyze the tweets, to find the correct meaning behind it whether it is offensive or not. In our project we are going to use various techniques such as Stop Words, Lexicon Analysis, Datasets and Machine Learning to Analyze
tweets and find out the sentiments behind it. We are also going to use Apache Spark based parallel processing technique to access only the latest tweets and not the old ones which are already being analyzed.

References

3. Hajime Watanabe, Mondher Bouazizi and Tomoaki Ohtsuki, Hate Speech on Twitter, IEEE, 2018.
12. J. Zhao, L. Dong, J. Wu, and K. Xu, Moodlens: An emoticon-based sentiment analysis system for Chinese is used.
13. A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, Learning word vectors for sentiment analysis, in Proc. ACL, 2011.

Index Terms

Computer Science
Information Sciences

Keywords

Twitter, Sentiment Analysis, Stop words, Lexicon Analysis, Machine Learning, Apache Spark