Chapter 2
Text Mining Using Twitter Data

Falak Bhardwaj
Manav Rachna International Institute of Research and Studies, India

Pulkit Arora
Manav Rachna International Institute of Research and Studies, India

Gaurav Agrawal
Manav Rachna International Institute of Research and Studies, India

ABSTRACT

The microblogging social networking service Twitter has been abuzz around the globe in the last decade. A number of allegations as well as exculpation of different types are being held against it. The list of pros and cons of social networks is huge. India on one hand had an abundance of internet access in last half of the decade. The growth of social media and its influence on people have affected the society in both good as well as in bad way. The following research was done in the month of September and October. The research was carried out on 13 lakh tweets approximately, collected over the course of a month from September to October providing insights about the different attributes of general tweets available on Twitter API for analysis. Insights include the hashtags, account mentions, sentiment, polarity, subject, and object of a tweet. The topics like Rhea Chakraborty and Sushant Singh Rajput, PM Narendra Modi's Birthday, IPL 2020 overshadowed the topics like COVID-19 and women's security.

INTRODUCTION

Twitter is a social networking service used for microblogging by more than 321 million active users with monetizable active accounts, worldwide. United States having the most, 68.7 million users and India being the third highest number of users at 18.9 million active accounts. Since its launch in 2006, twitter has been the source of information as well as misinformation. A large part of the microblogging service is used for news and announcements while in recent times a larger part of it is used to spread misinformation, fake news and to run a particular propaganda. Hence, it has affected major sectors like economy and politics, globally. In India, since the introduction of certain telecom providers in 2016,

the internet accessibility doubled up, which led to the exposure of the public to the internet. The sudden change and the digitalization escaped internet literacy. The abundant accessibility of the internet provided the end user with all types of information and data. Government bodies use twitter to provide important updates related to anything and everything while different bodies use it to provide users with different services. Celebrities and personalities use it to interact with their people. The significant use of the microblogging service is, it lets users share their views related to any social, economic, political or any other demographic topic with other people. The rapid increase in the usage of social networking websites provides an insight into many research challenges related to data mining and gained knowledge. Traditionally, the internet was comprehended as an information corpus, where users are passive. Social networking sites paved the way where users can create, publish and share intellectual contents online. It enhances the community strength and reach as people interact over a particular view related to a particular topic. It may lead to agreement or disagreement; in any case it ends up in interaction among people. As certain as it is that it leads to agreement and community strength, disagreement leads to riots and political imbalance. The internet literacy rate is almost proportional to the rate of increase of internet users in India. Social media has affected the youth as well as the elder generation of this time. The unemployment rate has led the youth to spend more time on social media. The hate speech content gets shared widely so easily due to a sole reason of communalism and fake news. The content once shared on social media cannot be retrieved back as the reach of such content is so wide and fast. In India, in recent times, there are more than 100 cases of riots and mob lynching registered due to social media and fake reports being generated over them. While the misinformation led to riots on ground level, the news media was involved in some cases. The misleading facts and the act of misleading is creating havoc among the people who are consumers of such media content. In the months of September and October the COVID-19 cases surged and the need to look over the situation was high. But, the topics of the conversation around the media do not look even around healthcare. Among social media, Twitter recently attracted researchers due to its sudden growth. The objective of sentiment analysis is to identify and extract sentiments from user-generated content. There has been a progressive shift in the area from review websites to micro blogs. Twitter sentiment analysis in itself is challenging due to its unique features such as the length of the tweet which is limited to 240 characters. Hence, the research was carried out using the Twitter API to extract real-time data available for free access for analysis at academic level. Then, the extracted data was analyzed and only the required fields and attributes were saved and later analyzed based on the attributes and requirements. This chapter focuses on the short sentences and entity level sentiment analysis and classifies the streamed tweets into positive, neutral and negative tweets using standard classifier. There are many challenges in Sentiment Analysis. Firstly, an opinion word which is considered to be positive in one state may be considered negative in a different situation. Second one, people may not always express opinions in a similar manner, e.g.: "the picture was a great one" differs completely from "the picture was not so great". The Opinions of people may be contradictory in their statements. It is more difficult for a machine to analyze. Most of the time people find it difficult to understand what others mean within a short sentence of text because it lacks context. Sentiment analysis is done on three levels:

1. Document Level: Analysis is done on the whole document and then expresses whether the document has positive or negative sentiment.
2. Sentence level: It is related to finding sentiment polarity from short sentences. Sentence level is merely close to subjectivity classification.

Related Content

Revolutionizing Conversational AI: Unleashing the Power of ChatGPT-Based Applications in Generative AI and Natural Language Processing

C. V. Suresh Babuand P. M. Akshara (2024). *Advanced Applications of Generative AI and Natural Language Processing Models (pp. 228-248).*

www.irma-international.org/chapter/revolutionizing-conversational-ai/335840

Emotion Recognition From Speech Using Perceptual Filter and Neural Network

Revathi A.and Sasikaladevi N. (2020). *Neural Networks for Natural Language Processing (pp. 78-91).*

www.irma-international.org/chapter/emotion-recognition-from-speech-using-perceptual-filter-and-neural-network/245084

Harnessing the Power of ChatGPT to Explore Student Metacognitive Skills in Learning Sociology Education

Ahmad M. Al Yakin, Ahmed J. Obaid, L. Abdul, Idi Warsah, Muthmainnah Muthmainnahand Ahmed A. Elngar (2024). *Advanced Applications of Generative AI and Natural Language Processing Models (pp. 405-423).*

www.irma-international.org/chapter/harnessing-the-power-of-chatgpt-to-explore-student-metacognitive-skills-in-learning-sociology-education/335849

Language Resource Acquisition for Low-Resource Languages in Digital Discourses

Kathiravan Pannerselvam, Saranya Rajiakodiand Bharathi Raja Chakravarthi (2024). *Empowering Low-Resource Languages With NLP Solutions (pp. 11-24).*

www.irma-international.org/chapter/language-resource-acquisition-for-low-resource-languages-in-digital-discourses/340499

Study of Neural Machine Translation With Long Short-Term Memory Techniques

Mangayarkarasi Ramaiah, Debajit Datta, Vanmathi C.and Rishav Agarwal (2023). *Deep Learning Research Applications for Natural Language Processing (pp. 65-88).*

www.irma-international.org/chapter/study-of-neural-machine-translation-with-long-short-term-memory-techniques/314136