



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 Issue: XI Month of publication: November 2019

DOI: <http://doi.org/10.22214/ijraset.2019.11012>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Real Time Twitter Trend Mining System

Miss. Priyanka Pawar¹, Miss. Ankita Deshmukh², Prof. P. A. Chougule³, Miss. Shital Shinde⁴, Mr. Utkarsh Swami⁵
^{1, 2, 3, 4, 5}Information Technology Department, Dr. J.J. Magdum College of Engineering, (Shivaji University, Kolhapur)

Abstract: *The advent of social media is changing the existing information behavior by letting users access to real-time online information channels without the constraints of time and space. It also generates a huge amount of data worth discovering novel knowledge. Social media, therefore, has created an enormous challenge for scientists trying to keep pace with developments in their field. Most of the previous studies have adopted broadbrush approaches which tend to result in providing limited analysis. To handle these problems properly, we introduce our real-time Twitter trend mining system, RT²M, which operates in real-time to process big stream datasets available on Twitter.*

Keywords: *Social media mining; real-time Twitter trend mining system; topic modeling; network analysis; Korean presidential election*

I. INTRODUCTION

Social media is changing the existing information behavior by letting users' access to real-time online information channels without the constraints of time and space. It also generates a huge amount of data worth discovering novel knowledge. Social media, therefore, has created an enormous challenge for scientists trying to keep pace with developments in their field. Most of the previous studies have adopted broad-brush approaches which tend to result in providing limited analysis. To handle these problems properly, we introduce our real-time Twitter trend mining system, RT²M, which operates in real-time to process big stream datasets available on Twitter. The system offers the functions of term co-occurrence retrieval, visualization of Twitter users by query, similarity calculation between two users, Topic Modeling to keep track of changes of topical trend, and analysis on mention-based user networks. Social media is a representative form of the Web 2.0 services which has undoubtedly contributed to the emergence of Big Data, based on voluntary and active participation of users. The most popular tools and frameworks for real time stream data processing in the Big-Data era will be presented. The paper presents the Big Data concept by defining it from different perspectives, the various processing models followed by an analysis of the currently industrial and research frameworks. The purpose of this paper is to present the necessary concepts regarding the different Big Data processing models, to propose and implement a system which make use of two new stream processing frameworks from Apache -Storm and Heron, in order to determine the most discussed subjects from the Twitter social network and their classification at the geolocation level.

II. DATASET

Analyzing structured data have been widely used. In such case, the traditional Relational Database Management System (RDBMS) can deal with the data. With the increasing amounts of unstructured data on various sources (e.g. Web, Social media, and Blog data) that are considered as Big Data, a single computer processor cannot process such huge amount of data. Hence, the RDBMS cannot deal with the unstructured data; a nontraditional database is needed to process the data, which is called NoSQL database. Most studies focused on tools, such as R (the programming language and the software environment for data analysis). R has limitations when processing twitter data, and is not efficient in dealing with large volume of data. To solve this problem a hybrid big data framework is usually employed, such as Apache Hadoop (an open source Java framework for processing and querying vast amounts of data on large clusters of commodity hardware) . Hadoop also deals with structured and semi-structured data, XML/JSON files, for example. The strength of using Hadoop comes in storing and processing large volume of data, while the strength of using R comes in analyzing the already-processed data.

III. RELATED WORK

Social media draws every attention from various scholastics communities for it offers a whole new opportunity it understands the social practices. diverse academic community, therefore, focus on various methodologies such as text mining network analysis, opinion mining, and etc. Our study aims at designing and developing the RT²M (Real-time Twitter Trend Mining) system which allows in real-time to Store every textual data ("tweets") generated in Twitter. Keep track of social issues by temporal Topic Modeling, and Visualize mention-based user networks.

Simply focuses on generic connectivity [5]. In other words, we concentrate on identifying thematic coherence among user's resident in the action of sending receiving mentions. Meanwhile search community's employee various approaches for mining unrevealed information from social data. In addition, SCAN algorithm suggests faster technique than other community's detection approaches

in identifying hub nodes connecting communities as well as removing outliers that plays a significant rule in the network. Modelling and community detection were devised for identifying relationship between search topic and scholarly communities' work also employs modularity and voltage clustering algorithm detect invisible community existing among users. A typology that includes the following four types: news, ongoing events, and memes. In other words, we concentrate on identifying thematic coherence among user's resident in the action of sending receiving mentions. Meanwhile search community's employee various approaches for mining unrevealed information from social data. In addition, SCAN algorithm suggests faster technique than other community's detection approaches in identifying hub nodes connecting communities as well as removing outliers that plays a significant rule in the network. Modelling and community detection were devised for identifying relationship between search topic and scholarly communities' work also employs modularity and voltage clustering algorithm detect invisible community existing among users. A typology that includes the following four types: news, ongoing events, and memes. Convolution, thus, regularly includes back propagation so as to all the more precisely weight the final result.

IV. METHODOLOGY

We introduce our twitter mining system which operates in real-time to process big stream datasets available on twitter.

- 1) *Term Co-occurrence Retrieval:* a query, the system retrieves the list of terms co-occurred with the query term. Once the list is obtained, it sorts co-occurred terms by co occurrence frequency and can display the result with an option of 100, 500, 1,000, and 2,000 terms. Like other functions provided in the system, co-occurred terms are dynamically updated and displayed as more Twitter stream data is received.
- 2) *Visualization of Twitter Users by Query:* Upon the query term, the system visualizes the social network graph of Twitter users mentioned together with the query term. The visualized graph is bidirectional. The social network analysis for the mention-based Twitter users is provided in the next section.
- 3) *Topic Modelling:* The system generates N topics specified by a user along with topical terms for the user selects. For the topic modelling technique, we employ the Multinomial Latent Dirichlet Allocation technique. If a start time and end time are entered, the system returns chronological topic trends related to queries.

V. RESULT

In In this Project, we are developing the Real-Time Twitter trend Mining (RT2M) system that is designed for in Real-Time to Store every tweet produced in tweet. Keep track of Topical trend Mine dynamic social trends and utilize them for betterment using R language.

VI. CONCLUSION

In this we study how to apply sentiment analysis to twitter data to observe changes in public opinion and the formation process of a certain issue, and ultimately design the prediction model of social issues on social media. In addition, by applying the concept of citation to Twitter, we may be able to discover content-based influencers and opinion leaders.

REFERENCES

- [1] Blei, D., Ng, A., and Jordan, M. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3: 993-1022.
- [2] Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K. P. 2010. Measuring user influence in Twitter: the million followers. *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*.
- [3] Chaney, A., Blei, D. 2012. Visualizing topic models. *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*.
- [4] Choi, D., Min, M., Kim J., and Lee, J. 2011. A Study on topic tracking using microblog. *Proceedings of KIIS Spring Conference 2011*.
- [5] Huberman, B. A., Romero, D. M., and Wu, F. 2008. Social networks that matter: Twitter under the microscope. Available at <http://ssrn.com/abstract=1313405>



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)