Divya Khullar el at. International Journal of Recent Research Aspects ISSN: 2349-7688, Vol. 3, Issue 3, September 2016, pp. 104-105

Review of Social Media Analysis by Machine Learning

Divya Khullar[1], Gagan Kumar[2]

[1]M.tech Scholar, Modern Institute of Engineering & Technology
[2]Assistant Professor, Modern Institute of Engineering & Technology

***Abstract*— Tweets have reported everything from daily life stories to latest local and worldwideevents. Twitter content reflects real-time events in our life and contains rich social information and temporal attributes. Monitoring and analyzing this rich and continuous flow ofuser-generated content can yield unprecedentedly valuable information. The popularity of microblogging stems from its distinctive communication services such as portability, immediacy, and ease of use, which allow users to instantly respond and spread information with limited or no restrictions on content. Virtually any person witnessing or involved in any event is nowadays able to disseminate real-time information, which can reach the other side of the world as the event unfolds. For instance, during recent social upheavals and crises, millions of people on the ground turned to Twitter to report and follow significant events.**

***Keywords*— Tweets, machine learning, data mining, clustering, tf-idf**

I. INTRODUCTION

Today, the textual data on the internet is growing rapidly. Several kinds of industries are trying to use this massive textual data for extracting the people's views towards their products. Social media is a crucial source of information in this case. It is not possible to manually investigate the heavy amount of data. This is where the requirement of automatic classification becomes clear. Subjective data is investigated commonly in this case [8]. There are a large number of social media websites that ensures users to supply modify and grade the content. Users have a freedom to express their personal views about specific topics. The example of such websites involves blogs, forums; product reviews sites, and social networks. In this case, twitter data is used. Sites like twitter consists of extensive

Tweets have reported everything from daily life stories to latest local and worldwideevents. Twitter content reflects real-time events in our life and contains rich social information and temporal attributes. Monitoring and analyzing this rich and continuous flow ofuser-generated content can yield unprecedentedly valuable information. The popularity of microblogging stems from its distinctive communication services such as portability, immediacy, and ease of use, which allow users to instantly respond and spread information with limited or no restrictions on content. Virtually any person witnessing or involved in any event is nowadays able to disseminate real-time information, which can reach the other side of the world as the event unfolds. For instance, during recent social upheavals and crises, millions of people on the ground turned to Twitter to report and follow significant events.

Unlike other media sources, Twitter messages provide timely and fine-grained information about any kind of event, reflecting, for instance, personal perspectives, social information, conversational aspects, emotional reactions, and controversial opinions.

Twitter has evolved over time and adopted suggestions originally proposed by users to make the platform more flexible. It currently provides different ways for users to converse and interact by referencing each other in posted messages in a well-defined markup vocabulary. Twitter is becoming the microphone of the masses, which altered news production and consumption (Murthy 2011). Many real-world examples have shown the effectiveness and the timely information reported by Twitter during disasters and social movements. Representative examples include the bomb blasts in Mumbai in November 2008, the flooding of the Red River Valley in the United States and Canada in March and April2009, the U.S. Airways plane crash on the Hudson river in January2009, the devastating earthquake in Haiti in 2010, the demonstrations following the Iranian Presidential elections in 2009, and the "Arab Spring" in the Middle East and North Africa region.

Several studies have analyzed Twitter's user intentions (Java et al. 2007; Krishnamurthyet al. 2008; Zhao and Rosson 2009; Kwak et al. 2010; Kaplan and Haenlein 2011). For instance, Java et al. (2007) categorized user intentions on Twitter into daily chatter, conversations, sharing information, and reporting news. They also identified Twitter users as information sources, friends, and information seekers. Krishnamurthy et al. (2008) presented similar classification of user intentions and also included evangelists and spammers that are looking to follow anyone. According to Kaplan and Haenlein (2011), people are motivated by the concept of ambient awareness—being updated about even the most trivial matters in other peoples' lives and by the platform for virtual exhibitionism and voyeurism provided for both active contributors and passive observers.

II. LITERATURE REVIEW

Arantxa Barrachina Arantxa Duque et.al. [1]: proposed. Technical Support call centres frequently receive several thousand customer queries on a daily basis. Traditionally, such organisations discard data related to customer enquiries within a relatively short period of time due to limited storage capacity. However, in recent years, the value of retaining and analyzing this information has become clear, enabling call centres to identify customer patterns, improve first call resolution and maximise daily closure rates. This paper proposes a Proof of Concept (PoC)

end to end solution that utilises the Hadoop programming model, extended ecosystem and the Mahout Big Data Analytics library for categorising similar support calls for large technical support data sets. The proposed solution is evaluated on a VMware technical support dataset.

Chen Min, et.al. [2]: They review the background and state-of-the-art of big data. They first introduce the general background of big data and review related technologies, such as could computing, Internet of Things, data centers, and Hadoop. Then focus on the four phases of the value chain of big data, i.e., data generation, data acquisition, data storage, and data analysis. For each phase, they introduce the general background, discuss the technical challenges, and review the latest advances. Finally examine the several representative applications of big data, including enterprise management, Internet of Things, online social networks, medial applications, collective intelligence, and smart grid. These discussions aim to provide a comprehensive overview and big-picture to readers of this exciting area. This survey is concluded with a discussion of open problems and future direction.

IoannisPartalas et al[4]: This paper provides an overview of the workshop Web-Scale Classification: Web Classification in the Big Data Era which was held in New York City, on February 28th as a workshop of the seventh International Conference on Web Search and Data Mining. The goal of the workshop was to discuss and assess recent research focusing on classification and mining in Web-scale category systems. The workshop brought to-gether members of several communities such web mining, machine learning, text classification and social media mining.

Lu Guofan, et.al. [7]: For call tracking system to adapt to the needs of large data processing, combined with a strong competitive advantage in recent years in large data processing Hadoop platform, designed and implemented a Hadoop-based call tracking data processing model, in order to verify its feasibility. The call tracking processing system model contains an analog data source module, data processing module, and a GUI interface. Analog data source module from real data samples in the simulated data, and the data is written directly to the Hadoop distributed file system, then using Hadoop's MapReduce model to write appropriate Mapper and Reducer function, the distributed processing of the data. Detailed study based on the system design and implementation, system deployment topology, hardware and software conditions, and designed several comparative experiments to analyze some static indicators of system performance.

Min Chen et al[8]: In this paper, we review the background and state-of-the-art of big data. We first introduce the general background of big data and review related technologies, such as could computing, Internet of Things, data centers, and Hadoop. We then focus on the four phases of the value chain of big data, i.e., data generation, data acquisition, data storage, and data analysis. These discussions aim to provide a comprehensive overview and big-picture to readers of this exciting area. This survey is concluded with a discussion of open problems and future directions.

III. CONCLUSIONS

The call tracking processing system model contains an analog data source module, data processing module, and a gui interface. Analog data source module from real data samples in the simulated data, and the data is written directly to the hadoop distributed file system, then using hadoop's mapreduce model to write appropriate mapper and reducer function, the distributed processing of the data. Detailed study based on the system design and implementation, system deployment topology, hardware and software conditions, and designed several comparative experiments to analyze some static indicators of system performance.

IV. REFERENCES

[1]. Arantxa Duque Barrachina, Aisling O'Driscoll. A big data methodology for categorising technical support requests using Hadoop and Mahout .Journal of data 2014: doi: 10.1186/2196-1115-1

[2]. Chen, Min, Shiwen Mao, and Yunhao Liu. "Big data: a survey." *Mobile Networks and Applications* 19.2 (2014): 171-209.

[3]. IoannisPartalas,, et al. "Web-scale classification: web classification in the big data era." *Proceedings of the 7th ACM international conference on Web search and data mining*.ACM, 2014.

[4]. Lu GuofanQingnian Zhang, Zhao Chen. Telecom Data processing and analysis based on Hadoop. Received 1 October 2014: Computer Modeling & New Technologies 2014 18(12B) 658-664.

[5]. Min Chen, Shiwen Mao, Yunhao Liu. Big Data: A Survey: Science+Business Media New York 2014. Springer Mobile NetwAppl (2014) 19:171–209 .DOI 10.1007/s11036-013-0489-0.

[6]. Scott Monteith, Tasha Glenn, John.Big data are coming to psychiatry. Monteith et al. Int J Bipolar Disord (2015) 3:21DOI 10.1186/s40345-015-0038-9.

[7]. Olshannikova Ekaterina, Aleksandr Ometov1, Yevgeni. Visualizing Big Data with augmented and virtual reality: challenges and research agenda.Olshannikova et al. Journal of Big Data (2015) 2:22: DOI 10.1186/s40537-015-0031

[8]. Sunil B. Mane, YashwantSawant, SaifKazi, VaibhavShinde, 'Real Time Sentiment Analysis of Twitter Data Using Hadoop', Vol. 5 (3) , 2014, 3098 - 3100 3098

[9]. T.K.Das and P.Mohan Kumar, 'BIG Data Analytics: A Framework for Unstructured Data Analysis' , Vol 5, No1, Feb-Mar 2013

