

Novel Method of Analysis of Tweets by Naive Bayes with Ant Colony Optimization

Divya Khullar¹, Gagan Kumar²

¹M.tech Scholar, Modern Institute of Engineering & Technology

²Assistant Professor, Modern Institute of Engineering & Technology

Abstract— Practically any person witnessing or concerned in any event is nowadays able to disseminate real-time information, which can reach the other side of the world as the occurrence unfolds. For instance, during recent social upheavals and crises, millions of people on the ground twisted to Twitter to report and follow significant events. Online social media sites (Facebook, Twitter, Youtube, etc.) have revolutionized the way we communicate with individuals, groups, and communities and altered every day practices. Twitter provides an application programming interface (API),⁵ which allows developers to programmatically access the public data streams as well as many features of the service. In this thesis proposed approach reducing classification by ant colony optimization, which reduce the features error.

Keywords— features,ACO,naivebayes ,classifier

I. INTRODUCTION

Mainly general definition explains characteristics of big data as volume, velocity and variety. Volume refers to the massive size of big datasets. Velocities refers to the rate at which data are generated and must be acted upon, such as filtered, decrease, transferred and examined, as opposed to stored for future processing [7]. Today, the textual data on the internet is growing rapidly. Several kinds of industries are trying to use this massive textual data for extorting 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. 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.

1.1 Tweets

The reputation of microblogging stems from its unique communication services such as portability, immediacy, and ease of use, which permit users to directly respond and spread information with limited or no restrictions on content. Practically any person witnessing or concerned in any event is nowadays able to disseminate real-time information, which can reach the other side of the world as the occurrence unfolds. For instance, during recent social upheavals and crises, millions of people on the ground twisted to Twitter to report and follow significant events. Online social media sites (Facebook, Twitter, Youtube, etc.) have revolutionized the way we communicate with individuals, groups, and communities and altered everyday practices (Boyd and Ellison 2007). Twitter provides an application programming interface (API),⁵ which allows developersto programmatically access the public data streams as well as many features of the service.For instance, Twitter streaming API provides filtering by location, keywords, author, and others. The availability of Twitter data has stimulated important research work in different disciplines and led to numerous applications and tools.

1.2 Twitter as a Source of Information

Twitter is flattering the microphone of the masses, which altered news making and consumption (Murthy 2011). Numerous real-world examples have shown the efficiency and the suitably information reported by Twitter at the time of disasters and social movements. Representative examples consist of the bomb blasts in Mumbai in November 2008, the

flooding of the Red River Valley in the United States and Canada in March and April 2009, the U.S. Airways plane crash on the Hudson river in January 2009, the distressing 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.

Numerous studies have examined Twitter's user purposes (Java et al. 2007; Krishnamurthy et al. 2008; Zhao and Rosson 2009; Kwak et al. 2010; Kaplan and Haenlein 2011). For instance, Java et al. (2007) classified user intentions on Twitter into daily chatter, conversations, sharing information, and reporting news. They also identified Twitter users as information resources, 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 powerful exhibitionism and voyeurism provided for both active contributors and passive observers.

II. LITERATURE REVIEW

Arantxa Barrachina Arantxa Duque et.al [1]: author proposed Technical Support call centres which oftenly receive thousand's of customer queries on daily basis. By custom, such type of organisations reject data which are connected to customer enquiries within a somewhat short span of time because of limited storage ability. While, in current years, the importance of obtaining and regulating this information has become very apparent, enabling call centres to be acquainted with customer paradigm, improved prime call declaration and increases everyday finish rates. This paper proposes a Proof of Concept (PoC) end to end solution that employs the Hadoop programming model, wide-ranging ecosystem and the Mahout Big Data Analytics library for classifying associated support calls for immense technical support data sets. The predictable outcome is developed on a VMware technical support dataset.

Chen Min, et.al. [2]: They assessed the environment and state-of-the-art of big data. They major started the general background of big data and analyse interrelated technologies, alike as could computing, Internet of Things, data centres, and Hadoop. They then focused on the four phases of the value chain of big data, i.e., data generation, data attainment, data storage, and data examine. For individual phase, they originate

the common background, argued the technical problems, and evaluate the most recent improvements lastly, examined the several delegate uses of big data, involving enterprise management, Internet of Things, online social networks, medical applications, collective intelligence, and smart grid. These deliberations main motive is to supply a extensive sketch and huge picture to readers of this stimulating area.

Ioannis Partalas et al[3]: delivers an overview of the workshop Web-Scale is categorized: Web classification in the Big Data Era which was taken place in New York City, on February 28th as a workshop of the seventh International Conference on Web Search and Data Mining. The main aims of the workshop were to argued and assessed existed research concentration on classification and mining in Web-scale category framework. The workshop incorporate members of various communities alike web mining, machine learning, context categorization and social media mining.

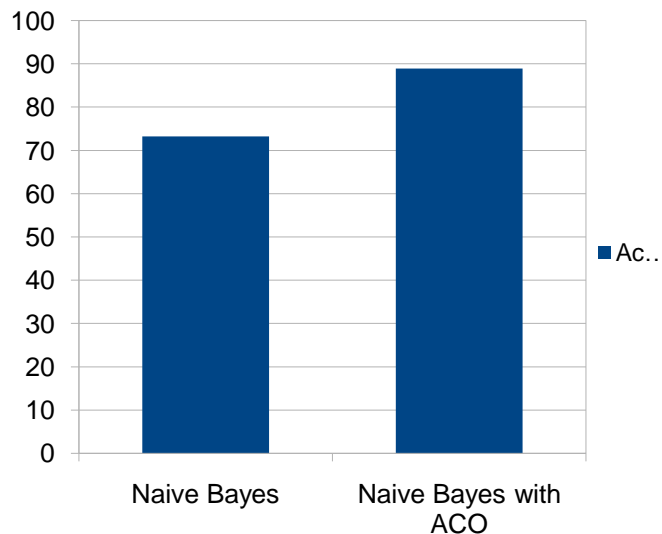
Lu Guofan, et.al. [4]: this paper presented call tracking processing system model composed of an analog data source module, data processing module, and a GUI interface. Analog data on the basis of component from genuine data samples in the pretend data, and the data is written directly to the Hadoop distributed file system, then with the use of Hadoop's MapReduce model to write appropriate Mapper and Reducer function, the discrete dispensation of the data. Inclusive research on the basis of system design and completion, system operation topology, hardware and software circumstances, and planned various relative experiments to scrutinize some stationary indicators of structure arrangement.

Min Chen et al[5]: analyzed the background and state-of-the-art of big data. They primarily launched the common background of big data and appraise interrelated technologies, alike could computing, Internet of Things, data centers, and Hadoop. They then focused 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 inclusive summary and big-picture to readers of this exciting area. This analysis is proficient with a conversation of open troubles and future directions.

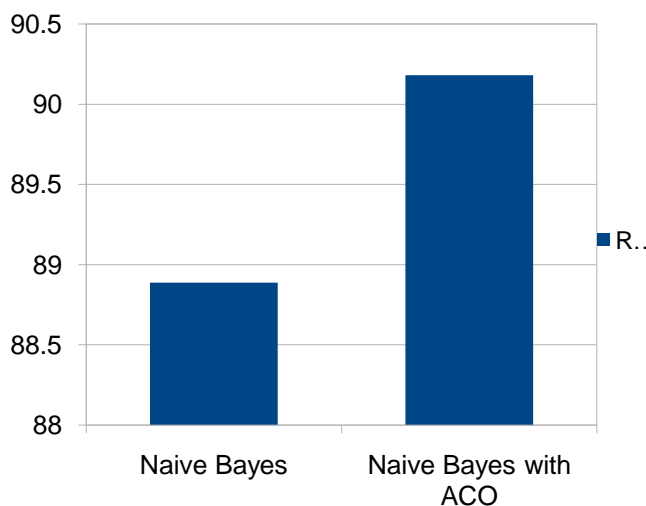
III. PROPOSED WORK

- Step1: Input the tweet text by continue streaming of tweets
- Step2: Preprocessing the text by tokenization stemming and stop-word removal
- Step3: Make vector space model with help of TF-IDF (Inverse document frequency)
- Step6: Hybrid the metaheuristic with classifier and make the classifier model
- Step7: Check the performance of classifier model by precision, recall and accuracy.

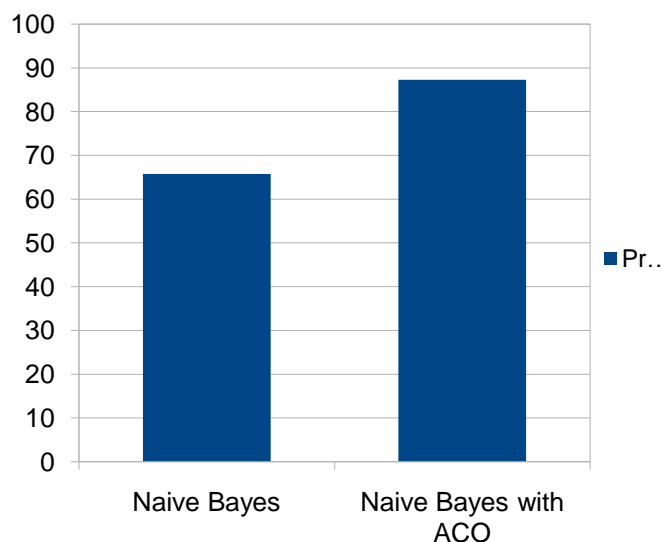
IV. Result Analysis



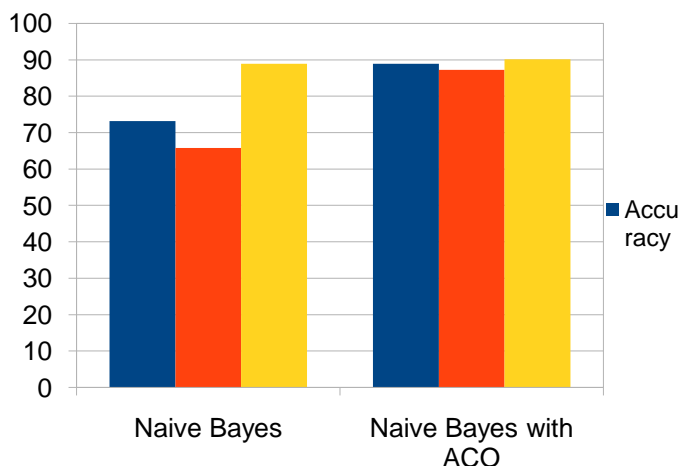
Graph 1: Comparison of Accuracy



Graph 2: Comparison of Recall



Graph 3: Comparison of Precision



Graph 4: Comparison between Recall, Precision and Accuracy

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