

K. Sunil Manohar Reddy
Department of CSE
Matrusri Engineering College
Hyderabad, India

Keywords: Big data, Sentiment Analysis, Volume, Velocity, Opinion Mining

any organization will need to rely on technologies that enable a scalable, accurate, and powerful analysis of these formats [2]. New cost effective solutions, such as Hadoop are changing this and allowing data of high volume, velocity, and variety to be much more easily analyzed. Hadoop is a massively parallel technology designed to be cost effective by running on commodity hardware. Today businesses can use Microsoft's Hadoop implementation, HDInsight, and SQL Server 2012 to effectively understand and analyze unstructured data, such as social media feeds, alongside existing Key Performance Indicators [3].

Social Big Data represents the data gathered from blogs and social media websites. These data are large in volume with noise and have the dynamic aspect (the frequent changes per day). They are recognized by a set of links (due to relationship between users), a non structural nature and lack of completeness. [4] These characteristics of social data make it different from other data on which data mining techniques are applied. The most common use of social media analytics is to mine customer sentiment in order to support marketing and customer service activities.

Social media mining is the process of representing, analyzing, and extracting actionable patterns from social media data. Social media mining introduces basic concepts and principal algorithms suitable for investigating massive social media data [5]. The six most representative research issues in the social media mining: influence propagation, expert finding, recommender systems, link prediction, community detection, opinion mining, mood analysis, prediction of trust and distrust among individuals.

Emotions are a mandatory part of human nature that can be considered as hereditary. Also it has been found that expression of a particular emotion by different human being is identical. Some persistent emotions that last much longer result in mood. Mood can be a result of a combination of certain emotions of a person. On the whole emotions can be categorized into two: basic and complex. Basic emotions are joy, sadness, anger, fear, disgust and surprise. The complex emotions are a



Today's consumers are heavily involved in social media, with users having accounts on multiple social media services. Social media gives users a platform to communicate effectively with friends, family, and colleagues, and also gives them a platform to talk about their favorite (and least favorite brands). This “unstructured” conversation can give businesses valuable insight into how consumers perceive their brand, and allow them to actively make business decisions to maintain their image to extract valuable business intelligence from Big Data,

combination of two or more basic emotions that are experienced by a person at an instance.

C. Opinion Mining

Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service.

Research work on sentiment analysis are based on three kinds of corpus: existing corpus used in previous work, manual corpus extracted from social media or automatic corpus extracted using APIs provided by different social media. The pre processing includes multiple task, such as breaking the text into sentences, elimination of stop words and N-gram processing and then sentiment classification.

III. DATA PROCESSING USING HADOOP

Social networks like Twitter and Facebook manage hundreds of millions of interactions each day. Because of this large volume of traffic, the first step in analyzing social media is to understand the scope of data needed to be collected for analysis. Quite often the data can be limited to certain hash tags, accounts, and key words [6].

Once the data collection is done, data can be loaded into Hadoop using familiar tools like SQL Server Integration Services, or purpose built tools like Apache Flume [7]. Data can often be gathered for free directly from a social media services public application interfaces, though sometimes there are limitations, or from an aggregation service, such as DataSift, which pulls many sources together into a standard format[8].

The next step is to transform data into a format that can be used for analysis. Data transformation in Hadoop is completed using a process called MapReduce. MapReduce jobs can be written in a number of programming languages, including. Net, Java, Python, and Ruby, or can be system generated by tools such as Hive (a SQL like language for Hadoop that many data analysts would be immediately comfortable with) or PIG (a procedural scripting language for Hadoop). MapReduce takes unstructured data and transform (map) it to something meaningful, and then aggregate (reduce) for reporting. All of this happens in parallel across all nodes in the Hadoop cluster.

The meaningful data obtained after MapReduce stored in Hadoop can be loaded into existing enterprise business intelligence (BI) platform or analyzed directly using powerful self-service tools like PowerPivot and PowerView. [9]Customers utilizing SQL Server as their enterprise BI platform have a variety of options to access their Hadoop data, including: Sqoop, SQL Server Integration Services, and Polybase (SQL Server PDW 2012 only).

Having social media data loaded into an existing enterprise BI platform allows dashboards to be created that give at glance information on how customers feel about a brand. This gives the ability to visualize how customer sentiment is affecting top line sales over time! This type of powerful analysis allows businesses to have the insight needed to quickly adapt, and it's all made possible through Hadoop.

IV. CONCLUSION

To implement big data mining on social media tools like HADOOP can be used. Big data presents more opportunities for research and references in the public sector as well as in technical progress [10] [11]. The challenges in data analyzing can be overcome by capturing techniques in big data.

V. REFERENCES

- [1] B. Brown, M. Chui and J. Manyika, "Are you Ready for the era of 'Big Data'?" "McKinsey Quarterly, McKinsey Global Institute, October 2011.
- [2] Intel IT Center, "Peer Research: Big Data Analytics", Intel's IT Manager Survey on How Organizations Are Using Big Data, August 2012.
- [3] K. Bakshi, "Considerations for Big Data: Architecture and Approach", Aerospace Conference IEEE, Big Sky Montana, March 2012.
- [4] U. Fayyad, G. Piatetsky-Shapiro and P. Smyth, "From Data Mining to Knowledge Discovery in Databases", American Association for Artificial Intelligence, AI Magazine, Fall 1996, pp. 37- 54.
- [5] M. Smith, C. Szongott, B. Henne and G. Voigt, "Big Data Privacy Issues in Public Social Media", Digital Ecosystems Technologies (DEST), 6th IEEE International Conference on, Campione d'Italia, June 2012.
- [6] A. Vailaya, "What's All the Buzz Around "Big Data?"", IEEE Women in Engineering Magazine, December 2012, pp. 24-31.
- [7] Intel IT Center, "Planning Guide: Getting Started with Hadoop", Steps IT Managers Can Take to Move Forward with Big Data Analytics, June 2012.
- [8] C. Eaton, D. Deroos, T. Deutsch, G. Lapis and P.C. Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data, Mc Graw-Hill Companies, 978-0-07-179053-6, 2012
- [9] B.Gerhardt, K. Griffin and R. Klemann, "Unlocking Value in the Fragmented World of Big Data Analytics", Cisco Internet Business Solutions Group, June 2012.
- [10] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh and A.H. Byers, "Big data: The next frontier for innovation, competition, and productivity", McKinsey Global Institute, 2011.
- [11] S. Madden, "From Databases to Big Data", IEEE Internet Computing, June 2012, v.16, pp.4-6.