Exploring the potential of Social Media Data using Text Mining to augment Business Intelligence

Dr. Ananthi Sheshasaayee[1], R. Jayanthi[2]
[1]Research Supervisor, [2]Research Scholoar,
PG and Research Department of Computer Science & Application, Quaid-E- Millath Government College for Women (Autonomous), Chennai 600 002

Abstract: In recent years, social media has become world-wide famous and important for content sharing, social networking, etc., The contents generated from these websites remains largely unused. Social media contains text, images, audio, video, and so on. Social media data largely contains unstructured text. Foremost thing is to extract the information in the unstructured text. This paper presents the influence of social media data for research and how the content can be used to predict real-world decisions that enhance business intelligence, by applying the text mining methods.

Keywords: Social Media; Social Networking; Text Mining; Business Intelligence; Unstructured Text.

I. INTRODUCTION

With the growth of increasing availability of network bandwidth and Internet technology, social media has emerged in recent years. Social Media is a group of Internet-based applications that is built on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content [1].

Social media is a platform where people create content, share their opinions, views and ideas. Examples include Facebook, MySpace, Digg, Twitter and JISC listservs on the academic side, because of its ease of use, speed and reach, social media is fast changing the public chat in society and setting trends and agendas in topics that range from the environment and politics to technology and the entertainment industry. Since social media can also be understood as a form of collective wisdom, it can be used to predict real-world outcomes. "The Wisdom of Crowds" is about the aggregation of information in groups, resulting in decisions that are often better than could have been made by any single member of the group [2]. The popular saying on social media goes as follows: "We use Facebook to schedule the protests, Twitter to coordinate, and YouTube to tell the world."[3]

In the case of social media, the enormity and high variance of the information that propagates through large user communities presents an interesting opportunity for

harnessing that data into a form that allows for specific predictions about particular outcomes, without having to institute market mechanisms. One can also build models to aggregate the opinions of the collective population and gain useful insights into their behavior, while predicting future trends. Moreover, gathering information on how people converse regarding particular products can be helpful when designing marketing and advertising campaigns [1], [4].

Mining the existing data within social media applications is promising to produce the required information to meet the relevant objective. However, social media data are vast, noisy, unstructured, and dynamic in nature, and thus mining the data is not easy to do. Section 2 of this paper explores the potential of social media and Section 3 addresses text mining in discovering more information as a best tool to mine social media data. Section 4 explains how intelligence is gained from the mined data, applied for wise decisions, followed by Section 5 with a Case study.

II. INFLUENCE OF SOCIAL MEDIA DATA FOR RESEARCH

Kietzmann, Hermkens et al. (2011) present a framework that defines social media by using seven functional building blocks: identity, conversations, sharing, presence, relationships, reputation and groups [5]. Social media changes the way of our life, users can not only connect with each other but also can create and share content. So far, social media analysis has become a critical problem for

both academia and industry. The explosion of user generated content on social media provides a wealth of opportunities for tapping into user preferences, assessments, opinions, reviews and ratings about content, products, increasingly popular among corporations and has been deployed at companies like HP, GE, Motorola, and in applications such as forecasting sales and project-completion times [6],[7].

Generally, organizations want to use the social media data to understand the needs and behavior of their customers or specific targeted groups of individuals with respect to the organizations' current or future products or services. There are three major approaches to looking at social media – channel reporting tools, overview score-carding systems and predictive analytic techniques (primarily text mining) [4].

Since the data is publicly available, sentiment analysis and influencers can be combined to prove that participants who are very negative in their sentiment are actually not highly regarded as thought leaders by the rest of the community. This goes against the popular marketing motto that negative users have a very high effect on the community at large. Sentiment analysis and opinion mining tools allow businesses to understand product sentiments, brand perception, new product perception, and reputation management [7].

III. TEXT MINING METHODS

In social networking websites, people generally use unstructured or semi-structured language for communication. In everyday life conversation, people do not care about the spellings and accurate grammatical construction of a sentence that may lead to different types of ambiguities, such as lexical, syntactic, and semantic (Sorensen 2009). There are multiple sources of data for businesses in the form of web pages, emails, video and image files, news and reports which are called semi structured or unstructured data. They are in different formats and searching, analyzing such data is a time consuming activity and may lead to poor decisions. To solve this problem, application of techniques of text mining for information extraction, automatic categorization or summarization of documents has become mandatory [8]. Therefore, extracting logical patterns with accurate information from such unstructured form is a critical task to perform. Text mining is a knowledge discovery technique that provides computational intelligence [9]

Bollier (2009) has identified the currently most fertile Contexts for text mining as politics, public health, and business [10].

Text mining is the process of deriving novel information from a collection of texts (also known as a corpus). By "novel information," we mean associations, hypotheses, or trends that are not explicitly present in the text sources being analyzed.

Large-scale deployments of text mining can require a significant amount of development work for specific applications. We may need to build a custom ontology (i.e. special terminology "dictionary" for the subject domain), develop business-rules based parsing engines, etc.

The concept of text mining may seem to be complicated, but understanding the process is easy if the task is broken down step by step. Using a simplified approach that breaks down the overall process into steps offers insight into how text mining works. The Main components of Text Mining systems are preprocessing tasks, core mining operations, presentation layer components and Refinement techniques [11]. **Preprocessing tasks** generally convert the information from each original data source into a suitable format before applying various types of feature extraction methods. **Core Mining Operations** are the heart of a text mining system and include pattern discovery, trend analysis and knowledge discovery algorithms. **Presentation layer components** include visualization tools and user-facing query editors. Finally, **Refinement techniques** include methods that filter out redundant information.

IV. BUSINESS INTELLIGENCE

Business intelligence (BI) systems combine operational data with analytical tools to present complex and competitive information to planners and decision makers, in order to improve the timeliness and quality of the decision-making process. Business intelligence systems provide actionable information delivered at the right time when decisions need to be made [12], [13], [14].

Mostly, Business Intelligence techniques are computer based, they help in analyzing past, current and future trends of the enterprise. Analyzing the data plays a major role in decision support system, query and reporting, data mining, complex event processing, online analytical processing, process mining, business performance management, text mining, statistical and predictive analysis.

An important aspect of Business Intelligence is knowledge management that helps companies in making good strategies through proper insight and experiences. BI helps companies to analyze their tremendous volumes of data for decision making but unfortunately not all social media data is structured and simple to understand as some data exists in unstructured or semi-structured form which results in time consuming search and interpretation.

Automatic detection of emotions in texts is becoming increasingly important from an applicative point of view.

Survey, blogs and review site are used to collect customer opinion about products to get knowledge about the reputation of the company in the market. Companies are interested to know about the people demand. These surveys are to be then summarized to produce a report about the good and bad aspects of particular products. The summary reports are then to be used for decision making equally by manufacturer, customer and merchant. For business intelligence, it is useful to classify each opinion according to the aspect of the business or transaction e.g. product quality, ordering or credibility [15].

Some of the applications of Text Mining methods in Social Media to influence Decision making

- Using the chatter from Twitter.com to forecast box-office revenues for movies [16].

- Emerging geopolitical events using advanced social media analytics techniques [17].

- Mining Customer Feedbacks for Actionable Intelligence [18].

- Analyzing Health Questions in Social Media [19].

- Predicting Oscar Award Winners.

- Survey that deals with the problem of evaluating the submissions to crowdsourcing websites on which data is increasing rapidly in both volume and complexity[20].

- Mining social media for customer satisfaction and how social media can be very useful to increase customer satisfaction in telecommunication industry [21].

- Hotel ranking system based on the average utility gain a consumer receives from staying in a particular hotel. By doing so, we can provide customers with the "best-value" hotels early on.

- Analyzing insurance Twitter posts. The results of these analyses help identify keywords and concepts in the social media data, and can facilitate the application of this information by insurers [22].

V. CASE STUDY

An example, tweets are collected from the Twitter for a time period about the cricket players "sachin tendulkar" and "virat kohli" which is considered as the "Bag of Words". The study was made with an objective of extracting feature words or patterns such as "Good Batsman", "Excellent Batsman", "Poor Bowler" etc and rank the players based on the feature words contained in the tweets. The main components of Text Mining applied for this study are as follows:

1) **Preprocessing tasks** includes all those routines, processes and methods required to prepare data for a text mining system's core knowledge discovery operations. Pre-processing includes Sentence splitting i.e identifying the sentence boundaries, tokenization i.e splitting a tweet given in Twitter into tokens, eg., separating words, punctuation, special characters, apostrophes etc.

2) **Core Mining Operations** are the text mining system and include pattern discovery i.e., in this case of analysing "tweets", the bag of words "sachin", "tendulkar", "virat", "kohli" are considered as the feature word extraction. Totally 58 tweets has been taken for this case study. This is considered as the text document contains 7935 words. After preprocessing, the document words are reduced to 612 words using feature word extraction. The term frequency tf of "sachin tendulkar" is 8.0065 and "virat kohli is 12.418. It is justified that virat kohli is considered to be the best upcoming batsman at that particular period.

3) **Presentation layer** components include visualization using a Frequency Distribution diagram implemented using Python is shown in the following figures. Figure1 represents a Frequency Distribution diagram for the player1 Sachin Tendulkar and Figure2 represents a Frequency Distribution diagram for the player2 Virat kohli.
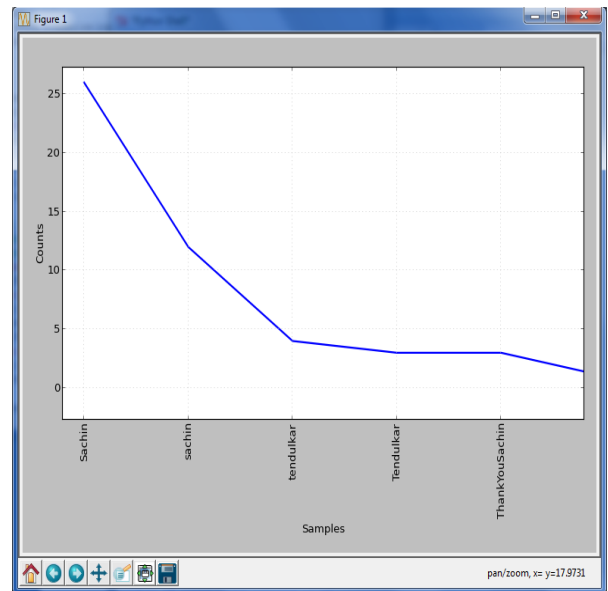

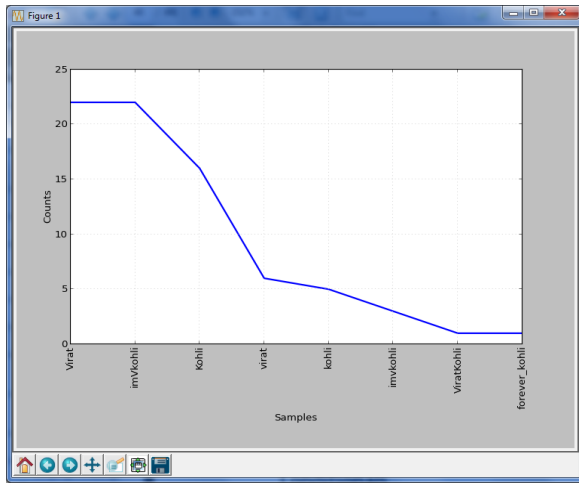Fig 1- Frequency Distribution diagram for Player1

Fig 2- Frequency Distribution diagram for Player2

4) **Refinement techniques** include methods that filter out redundant information. In this case study, the similar name to player2 like "forever_kohli " can be removed as it is not related to virat kohli. This is an example of a domain-specific refinement that has to be carried out to enhance the accuracy of the results of the mining operations.

This study was conducted for the social media data collected over a short period of time. This can be extended to large set of data and intelligence can be applied to enhance the decision making.

VI. CONCLUSION

This paper reviews some basic terminologies, definitions and applications of Social Media data using Text Mining methods and Business Intelligence. Some of the applications of Text Mining methods in Social Media which influences Decision making is also collected and discussed in this paper. A case study was considered from the twitter data and an analysis showed that which player is considered to be the best upcoming batsman for a particular period. Next focus of research would be to work on how Text Mining assists in extracting actionable intelligence from voluminous amounts of unstructured information sources of any social data within a domain system. Forward thinking organizations are now at the threshold of identifying how text mining fits within their current business intelligence framework and it is useful to classify each opinion according to the aspect of the business or transaction.

REFERENCES

[1] Zhu Aihua,Chen Xi "A review of social media and social business", Fourth International Conference on Multimedia Information Networking and Security, IEEE Computer Society 2012.

[2] James Surowiecki "Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations", 2005

[3] Anonymous Cairo activist, quoted in Nadine Kassem Chebib & Rabia Minatullah Sohail, "The Reasons Social Media Contributed to the 2011 Egyptian Revolution," International Journal of Business Research and Management 3 (2011): 139

[4] Efthimios Bothos et al, "Using Social Media to Predict Future Events with Agent- Based Markets", IEEE Intelligent System 2010.

[5] Kietzmann, J. H., K. Hermkens, et al. (2011). "Social media? Get serious! Understanding the functional building blocks of social media." Business Horizons 54(3): 241-251.

[6] Killian Thiel et al, "Creating Usable Customer Intelligence from Social Media Data: Network Analytics meets Text Mining" KNIME 2012.

[7] Pritam Gundecha, Huan Liu, Mining Social Media: A Brief Introduction, 2012 INFORMS, ISBN 978-0-9843378-3-5 http://dx.doi.org/10.1287/educ.1120.0105

[8] Palak Gupta et al, Gian-Jyoti E-journal, volume 1 issue 2,2012, Role of Text Mining in Business Intelligence.

[9] Rizwanairfan and et al, A Survey on Text Mining in Social Networks, The Knowledge Engineering Review, Vol. 00:0, 1–24, 2004, Cambridge University Press

[10] Bollier,D. (2009) 'The Promise and Peril of Big Data. Paper presented at Extreme Inference: Implications of Data Intensive Advanced Correlation Techniques', The Eighteenth Annual Aspen Institute Roundtable on Information Technology, Aspen, Colarado, The Aspen Institute. Retrieved February 12th, 2012, from http://bollier.org/sites/default/files/aspen_reports/InfoTech09_0.pdf

[11] S.Murugesh, Role of preprocessing tasks in Text Mining for software requirements elicitation, International Journal Of Research In Advance Technology In Engineering (IJRATE) Volume 1, Special Issue, October 2013

[12] Negash, S. (2004). Business Intelligence - communications of the Association for Information Systems, 13, 177-195

[13] Koronios, A., & Yeoh, W. (2010) "Critical success factors for business intelligence systems" Journal of Computer Information Systems, 23-32.

[14] Li Gao et al, World Academy of Science, Engineering & Technology 2007, Powerful Tool to expand Business Intelligence – Text Mining.

[15] Khairullah Khan, Baharum B.Baharudin, Aurangzeb Khan, Fazal-e-Malik, Mining Opinion from Text Documents: A Survey, 2009 3rd IEEE International Conference on Digital Ecosystems and Technologies

[16] Jayanth Marasanapalle, Vignesh T.S, Praveen Kumar Srinivasan, Angshuman Saha, Business Intelligence from Twitter for the Television Media: A Case Study,2010, IEEE

[17] Dr. Hsinchun Chen, From Dark Web to GeoPolitical Web: Exploring the Value of Social Media Informatics, 2012 European Intelligence and Security Informatics Conference

[18] Lipika Dey, Sk Mirajul Haque, Nidhi Raj, 2010 "Mining Customer Feedbacks for Actionable Intelligence", IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology

[19] Sanghee Oh, Min Sook Park, Text Mining as a Method of Analyzing Health Questions in Social Q&A, ASIST 2013, November 1-6, 2013, Montreal, Quebec, Canada

[20] Thomas P. Walter, Andrea Back, A Text Mining Approach to Evaluate Submissions to Crowdsourcing Contests, 2013 46th Hawaii International Conference on System Sciences

[21] Vitri Tundjungsari , Business Intelligence with Social Media and Data Mining to Support Customer Satisfaction in Telecommunication Industry, International Journal of Computer

Science and Electronics Engineering (IJCSEE) Volume 1, Issue 1 (2013) ISSN 2320–4028 (Online)

[22] Roosevelt C. Mosley, Social Media Analytics: Data Mining Applied to Insurance Twitter Posts, Casualty Actuarial Society E-Forum, Winter 2012-Volume 2

[23] Anindya Ghose, Panagiotis G. Ipeirotis, Beibei Li, Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content, 2012 INFORMS, Vol. 31, No. 3, May–June 2012, pp. 493–520 ISSN 0732-2399 (print) . ISSN 1526-548X (online)

[24] Lawrence Ampofo, Simon Collister, Ben O'Loughlin, and Andrew Chadwick, "Text Mining and Social Media: When Quantitative Meets Qualitative, and Software Meets Humans", New Political Communication Unit Working Paper, October 2013

[25] Dashboard Insight Article: "Understanding the role of text mining within BI The 4th in a series " by Lyndsay Wise July 27, 2007.

[26] Adam Funk et al, Opinion Analysis for Business Intelligence Applications.

[27] Anil Kumar Pantangi et al,SAS Global Forum 2012 on Customer Intelligence, Classification of Customers Textual Responses via application of Topic Mining.

[28] L. Francis, M. Flynn. (2010), Text Mining Handbook-Casualty Actuarial Society. [online] Available: www.casact.org/pubs/forum/10spforum

[29] P. N. Tan, M. Steinbach, V. Kumar, Introduction to Data Mining. Pearson AddisonWesley: Boston, 2006.

[30] T. M. Mann. (2002), Visualization of Search Result from the WWW, [online] Available: citeseerx.ist.psu.edu