A Survey on Social Media Analytics using Text Mining

Supreetha Pai[1] and Sowmya S R[2]

[1-2] Asst. Professor, Department of Information Science DSATM, Bangalore

Email:[1] paisupreetha@gmail.com, [2] srs.is.08@gmail.com

Abstract—**In the few years social media analytics has gained a very important role in people's life.Text mining techniques enable the user to identify patterns and relationships which exist within a body of texts which would otherwise be extremely difficult or time-consuming to discover. There is wide variety of social media channels with huge amount of data available. The challenge comes in accessing that data and transforming it into something that is usable and actionable. Generally, organizations want to use the social media data to understand the needs and behaviour of their customers or specific targeted groups of individuals with respect to the organization's current or future products or services.. In this paper a survey on social media analysis using text mining is discussed.**

Index Terms— **Social media, Twitter, analytics, opinion mining, Sentiment Analysis,Visual analytics.**

I. INTRODUCTION

Social media analytics is concerned with developing and evaluating informatics tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data to facilitate conversations and interactions.As the number of users on social media sites continues to increase, the need for businesses to monitor and utilize these sites to their benefit also increases. The amount of information seen during a single day in social networks gives the indication of social media's enormous influence. [Facebook has more than one billion active users, YouTube has over 4 billion views]. Facebook, Twitter integrates social actions in its online ads,allowing the users to see if their friends have liked or voted on products being advertised.

Social networking sites provides the user another quick and easy way to express their likes, dislikes, interests, and concerns, and these present further opportunities (or challenges) to businesses that want to stay abreast of these sentiments. There are a number of types of software tools for analyzing unstructured data found in tweets and Facebook posts. In addition to text analysis, many enterprise-level social media tools will harvest and store the data.

In this paper, how the explosion in social media necessitates the use of social media analytics is presented; The underlying stages of the social media analytics process is explained; the most common social media analytic techniques in useand the ways in which social media analytics create business value is discussed.

Social media analytics is a six step iterative process with following steps.
1. Data Identification: Searching and identifying the source information for analytical purposes.
2. Extraction: Once a reliable source of social media data are identified, data is extracted using API or manually.
3. Cleaning: This step involves removing the unwanted data from the extracted data
4. Analyzing: The cleaned data is analyzed for business insights.
5. Visualization: The analysis part will lead to relevant visualizations for effective communication
6. Interpretation: This step relies on human judgments to interpret valuable knowledge from the visual data.

II. SYSTEM MODEL

The data produced from blogs, discussion forums, social media and social networking sites, reviews are in the form of unstructured text. It is suggested text mining techniques are the way to pre-process and extract information from the unstructured text to make better business decision making.
It presents that identity, sharing, conversations, reputation, groups, relationships and presence are the seven building blocks of social media. it is opinioned that corporate makes use of social media that offers abundant occasions for gathering user preferences, opinions, ratings about a product or service and assessments.
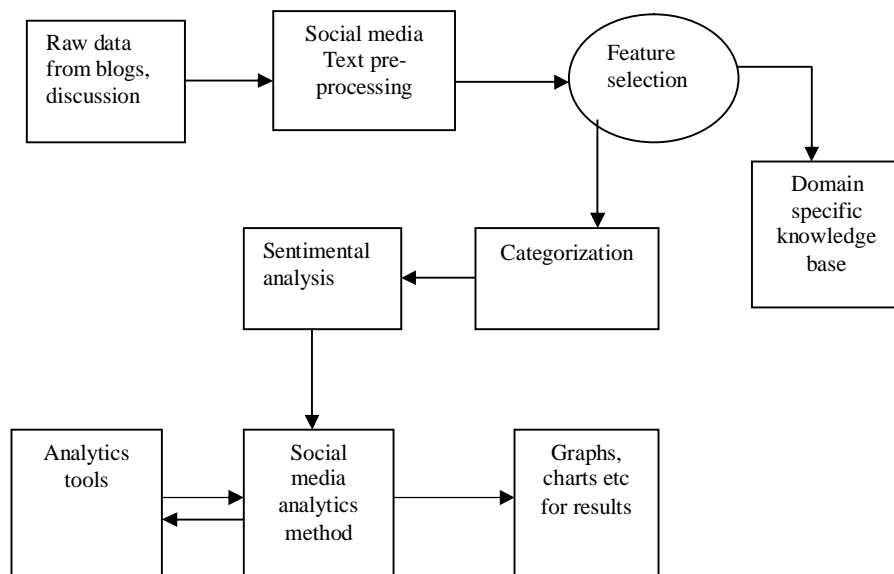

Figure 1: System model

It is suggested that the sentiments attached with a product, the perception about a brand and the perception about new product introduction can be well interpreted using sentiment analysis. According tosentiment analysis with the help of fuzzy logic deals with reasoning and gives closer views to the exact sentiment values. Accordingly an opinion mining system can be used for both binary and fine-grained sentiment classifications of user reviews. Feature-based sentiment classification is a multistep process that involves pre-processing to remove noise, extraction of features and corresponding descriptors and tagging their property using fuzzy functions.

III. TYPES OF ANALYSIS TOOLS

To track social media activities, following different types of tools are available.

A. Platform Analytics

There are some analytics tools provided by each of the platforms. For example, twitter analytics or Face book Insights.

B. Management Tool Analytics

There are wide varieties of social media management providers who provide analytics as part of their solution.

C. Standalone Analytics

These are social media measurement tools that only provide social media analytics functionality
Tools for Social Media Analysis

IV. TOOLS FOR SOCIAL MEDIA ANALYSIS

Some of the most commonly used Analytical tools are

A. *Keyhole*

It is used to measurein precise, a brand or trend's impact on Twitter, Facebook and Instagram. Giving access to an intuitive and shareable dashboard, it tracks hashtag, keyword and campaign metrics in real-time. These include reach, impressions, periods of high activity and more.

B. *Brandwatch*

It is used as a research suite. It provides information about the markets owned by user such as demographic data about gender and occupation.

C. *Little Bird*

Its goal is to eliminate the need for influencer research and instead allows to focus on outreach. The tool tracks metrics detailing the most prominent people who interact in user space.

D. *Google Analytics*

Itthe is use foranalyzing website traffic. And as part of the analysis, user can dig into the referral stats on social media marketing as wells.

E. *Quintly*

It covers Facebook, Twitter, Google+, LinkedIn, Inst1agram and YouTube, and it has a free tool for Facebook analytics. Quintly is a dashboard tool. It comes with a standard dashboard that can be customized with widgets to suit user needs and track the metrics

V. SOCIAL MEDIA ANALYTIC TECHNIQUES

Social media analytics is a growing area that encompasses a variety of modeling and analytical techniques from different fields. Here some techniques are explained below that are most instrumental in understanding, analyzing, and presenting large amounts of social media data. These techniques can support various stages of social media analytics.

1. *Sentiment analysis* and trend analysis primarily support the understand stage. Topic modeling and social network analysis have primarily applications in the understand stage but can support the capture and present stages as well. Visual analytics spans the understand and the present stages. Opinion mining (or sentiment analysis) is the core technique behind many social media monitoring systems and trend analysis applications. It leverage computational linguistics, natural language processing and other methods of text analytics to automatically extract user sentiments or opinions from text sources at any level of granularity. Such subjective information extracted about people, products, services, or other entities support various tasks including predicting stock market movements, determining market trends, analyzing product defects, and managing crises.

Though sentiment analysis is becoming more common, sampling biases in the data can badly skew results. There might be situations where satisfied customers remain silent while those with more extreme positions incessantly voice their opinions for large data.

2. *Topic modeling* is used for large bodies of captured text to detect dominant themes (topics). The themes uncovered can be used to provide consistent labels to explore the text collection further or to build effective navigational interfaces. Themes revealed by topic modeling can also be used to feed other analytical tasks such as discovering user interests, detecting emerging topics in forums or social media postings, or summarizing parts (or all) of a text collection.Topic modeling uses a variety of advanced statistics and machine learning techniques. For instance, a number of models identify "latent" topics by using the co-occurrence frequencies of words within a single communication or between topics and communities of users.

3. *Computational Science Techniques*is theautomated sentiment analysis of digital texts uses elements from machine learning such as latent semantic analysis, support vector machines, bag-of-words model and semantic orientation.

4. *Social network analysis* is used to analyze a social network graph to understand its underlying structure, connections, and theoretical properties as well as to identify the relative importance of different nodes within the network. A social network graph consists of nodes (users) and associated relationships (depicted by edges). The relationships are typically detected from user actions directly connecting two people (such as accepting another user as a "friend"), though they may be inferred from indirect behaviours creating relationships, such as voting, tagging, or commenting.

Social network analysis is the primary technique for identifying key influencers in viral marketing campaigns on Twitter or other social media platforms. It is used to detect sub-communities within a larger online community such as a discussion forum, allowing for greater precision in tailoring products and marketing materials. It has strong uses in predictive modeling, such as conducting marketing campaigns aimed at those assumed mostly likely to buy a particular product

Text mining and network mining in social media data as analytic approaches are widely used for new insights in social media data. In texts from comments or posts, blogs, the sentiments and opinions of users on certain topics, products, or persons are often mined. However, each technique follows its own specific goal.

VI. The Business Value of Social Media Analytics

There are a variety of business uses of social network. Some of them are discussed here.

A life cycle analysis perspective considers the life of a product (or service) from its design through its disposal, as well as support activities that take place in parallel with these activities. Here how social media analytics help firms gather competitive intelligence, over competitors is discussed.

Trend analysis and other social media analytic tools can help bring to the fore any changes in tastes, behaviours, and other sentiments that can affect product design and development. These tools can enable features to be added or adjusted, and they can help create sufficient lead time for creating "next generation" products or even products in a completely new category.

Social media analytics can also promote product innovation by capturing and understanding conversations involving either of two groups. The most advanced use of social media-based conversations is in the "co-creation" of products, where online users and businesses act as informal partners in generating new product ideas and even entirely new product categories. The risks during this life cycle phase involve supply chain responsiveness. Social media analytics can mitigate these risks.

Visual analytics can be useful in pointing out correlations, outliers, geographic patterns, or other trends that support smother functioning. Social media may give advance warning when situations become less predictable, including political tensions overseas that could disrupt the flow of metals, minerals, or other vital supplies for manufacturing.

VII. Conclusion

A set of complementary methodologies and visual analytical approaches are presented here. The social media data is usually unstructured. Hence to improve outcome from any of the methods, structuring of data is essential. Well structured data well lead to more precise parameter values for analytics. Few yet issues that should be explored Large-scale social media-based human behavior modeling, Text mining with humans in the loop, Intelligent Interactive systems.

REFERENCES

[1] Kaushik A.,Naithani K. ,A Comprehensive Study of Text Mining Approach .IJCSNS International Journal of Computer Science and Network Security (2016).

[2] Jan van Gemert, "Text Mining Tools on the Internet: An Overview". [11]E-book: Introduction to Data Mining andKnowledge Discovery Third Edition By Two Crows Corporation, ISBN: 1892095-02-5.

[3] Anderson, T. W. (1971). The statistical analysis of time series. New York: John Wiley & Sons.

[4] Asiedu, Y. and Gu, P. (1998). Product life cycle cost analysis: State of the art review. International Journal of Product Research.36(4).

[5] Billington, C., Lee H., and Tang C. (1998). Successful strategies for product rollovers. Sloan Management Review. 39(3).

[6] Blei, D. M. (2012). Probabilistic topic models. Communications of the ACM, 55(4).

[7] Bonchi, F., Castillo, C., Gionis, A., &Jaimes, A. (2011). Social network analysis and mining for business Applications. ACM Transactions on Intelligent Systems and Technology 2(3).

[8] Chen, H., Chiang, R. H. L., & Storey, V. (2012). Business intelligence and analytics: From big data to big impact. MIS Quarterly, 36(4).

[9] Chui, M., Manyika, J., Bughin, J., Dobbs, R., Roxburgh, C., Sarrazin, H, Westergren, M. (2012). The social economy: Unlocking value and productivity through social technologies.

[10] Turney, P. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised   classification of reviews. In Proceedings of the Association for Computational Linguistics. 417–424.

[11] Zeng, D., Chen, H., Lusch, R., & Li, S-H. (2010). Social media analytics and intelligence.
      IEEE   Intelligent Systems.

[12] S.-T. Wu, Y. Li, and Y. Xu, "Deploying Approaches for Pattern Refinement in Text Mining," Proc. IEEE Sixth Int'l Conf. Data Mining (ICDM '06), pp. 1157-1161, 2006.

[13] Divya Nasa, "Text Mining Techniques- A Survey" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 4, April 2012.

[14] E-book: Introduction to Data Mining and Knowledge Discovery Third Edition By Two Crows Corporation, ISBN: 1892095-02-5.

Supreetha Pai received her Bachelor of Engineering degree from Vivekananda College of Engineering and Technology ,Puttur in 2008.She received her Master of Technology degree in Digital Electronics from East west Institute of Technology, Bangalore in 2013. She is the gold medallist in M.Tech Digital Electronics examinations from VTU. Currently, she is an assistant professor in Information Science Department in Dayananda Sagar Academy of Technology, Bangalore.Her research interests are in Data analytics and image processing.

Sowmya S.R received her Bachelor of Engineering degree from Ghosia engineering College in 2008.She received her Master of Technology degree in Software engineering from AMC engineering College, Bangalore in 2012.Currently, she is an assistant professor in Information Science Department in Dayananda Sagar Academy of Technology, Bangalore.Her research interests are in Data analytics and image processing.