A Review on Social Media Analytics

Divya Kataria
Computer Science Engineering Department MUST, India
katariadivya94@gmail.com

Abstract: This paper highlights the wealth of social media analytics available. It provides a comprehensive review of the social networking media, wikis, blogs, newspaper group chat etc. Social media Analytics is an emerging interdisciplinary research field that aims at combining, extending and adapting methods for analytics of social media data. For completeness, it includes an introduction to analytics on data based on social media scraping storage, data cleaning and sentiment analytics. Major research area and business activities are based on analyzing social media, in particular, twitter feeds for sentiment analysis. Sentiment analysis for microblogs is built as standalone, endpoint to endpoint applications.

Keywords: Social Media, Challenges of Social Media Analytics, Text Analysis, Social Network Analytics, Trend Analysis, Software Platforms.

1. INTRODUCTION

In recent years, social media has a tremendous growth in user base. For example, there are one billion users of Facebook while twitter tends to have more than 280 million users. There are many social media applications available nowadays which can be categorized as weblogs, microblogs, social networking sites, wikis, location-based social networks, social bookmarking websites etc.

The paradigm shift in how people communicate, collaborate, create, and consume information is the mainstream adoption of social media applications. Specifically, the process of information consumption and dissemination is closely interrelated with the process of generating and sharing information (Zeng et al. 2010).

Everyday uses that are the source of social media's mass adoption is increasingly used as communication channels in the business, political, and other sectors like companies have started to adopt internal (private) as well as external (public) social media platforms for a number of purposes. Communication and collaboration among employees, knowledge management, and product/service innovation are improved with the use of internal social media. Various companies have started to establish social media based networks with business partners and have also begun to connect in public social media actions for the purposes of advertising, public relations, client relations, reputation administration, and employment. In the political sphere, social media is believed to have the potential to increase contribution by populace and electorate (Wattal et al. 2010) like Twitter is an ideal public platform for disseminating political information and opinions rapidly and broadly (Stieglitz and Dang Xuan 2013a, 2013b; Bruns and Highfield 2013), political actors like politicians, parties, fundamentals, etc. have also begun to use Facebook to enter into dialogues with citizens and to encourage more political deliberations.

Academic research from various disciplines of the social and even the natural and applied sciences has recently devoted more attention to social media. Social networks and social media have also become an important domain in information systems (IS) research. This recent significance in "Big Social Data" given by Manovich 2012; Burgess and Bruns in 2012 has been determined in part by facilitated right to use to large-scale pragmatic datasets from accepted online social networking platforms such as Twitter, Facebook, and LinkedIn, as well as from other platforms that facilitate group alliance and self-organization such as weblogs, wikis, and user tagging systems. Boyd and Crawford (2012) spotted out, "the era of Big Data is in progress. Computer scientists, physicists, economists, mathematicians, political scientists, bio-informaticists, sociologists, and other scholars are clamoring for access to the substantial quantities of information produced by and about the public, things, and their connections" (Boyd and Crawford 2012, p. 663). Defined as a cultural, scientifically, and scholarly happening that rests on the interaction of technology, analysis, and methodology, Big Data is "less about data that is big than it is about a capacity to search, aggregate, and cross reference large data sets" (Boyd and Crawford 2012, p. 663). From a research point of view, social media can be tacit as a kind of living wage lab, which enables academics to accumulate large amounts of data generated in real world surroundings.

There is a significant interest in analyzing Big Social Data from social media not only for research but also for practical purposes. For example, in analyzing social media data, companies see the opportunity for targeting advertising, PR, social customer

relationship management (CRM), and business intelligence (BI). In particular, the primary interest behind corporate activities in social media is how to effectively use them as an additional channel for marketing. Further, B2B companies have also started to use social media analytics to identify new potential customers. Recently, supporting institutions have also shown an oddity in monitoring the public view on policies and political positions, detecting trending political topics, and supervising their own reputation on the social web. Public officials could potentially use social media to identify situational information created by citizens in times of natural disasters (Bruns and Burgess 2012; Bruns and Liang 2012). Furthermore, based on social media data, health organizations could establish an early warning system for disease outbreaks that should help provide timely response measures. Also, individuals and consumers seek to make use of information and opinions from diverse sources in order to make more informed decisions. A further use case for research can be found in conducting intercultural studies: for example by analyzing social media content, academics are able to directly compare how people in different countries or cultures react to certain global events.

This interest has two main drivers. First, due to technological advances, there is a new possibility of continuous, automated (real time) monitoring, and analytics of social media content and interactions. Second, there is a change in public participation that leads to an increased complexity of the communication environment like increasing quantity and heterogeneity of communicators, unbounded communication, a higher level of information diffusion with respect to range, scale, and speed, in particular, due to the rapid development of mobile devices. In this sense, geo data are another highly promising data source which can be effectively included in social media analytics procedures.

1. SOCIAL MEDIA ANALYTICS AND INTERDISCIPLINARY METHODS

In the context of SMA, there are mainly three ways to tackle challenges related to Social Media data and find a better way for their applications: (1) text analysis/mining) (2) social network analysis (3) trend analysis.

Text analysis/mining is a research technique in the sphere of content analysis which supports researchers in driving replicable and valid inferences from texts to the contexts of their use (Krippendorff 2004). Based on this a large variety of queries can be solved, among which are the classification of texts (i.e., sentiment analysis) and the identification and modeling of recurring data (Krippendorff 2004). Recently, most notably text classification based on unsupervised and supervised learning (Sebastiani 2002; Liu 2011) came into being. Supervised text classification is based on statistical algorithms from machine learning such as support vector machine (SVM) or naive Bayesian classifier and this has become a standard method for automated text mining.

Text analysis/mining has one of its main component i.e. sentiment analysis or opinion mining, which has emerged as a distinct method to study people's opinions in terms of views, appraisals, emotions and attitude towards entities, individuals, issues, events, topics, and their attributes in a more thorough way (Pang and Lee 2008; Liu 2011).

The second main analysis method is social network analysis (SNA), which gives us the relationships between persons, organizations, interest groups, states, etc., by analyzing the structure of their connections (Scott and Carrington 2011). In an SMA context, SNA helps to identify influential users or opinion leaders, and relevant user communities in social media. There are different no of measures for the influence of an actor in a network. SNA, therefore, provides different metrics for the concept of centrality and prestige that can be applied to measure influence (Wasserman and Faust 1994; Scott and Carrington 2011). Regarding the detection of relevant communities, SNA can significantly help in different community detection methods and algorithms like graph-theoretical approaches such as the Girvan-Newman algorithm (Girvan and Newman 2002) or other clustering methods like fuzzy c-means clustering (Prabhu et al. 2010). However, SNA has many challenges relate to discovering changing clusters in large-scale and dynamic data.

Trend analysis is the third important analysis method which makes use of recent advances in computer science and statistics to predict emerging topics. Many trend detecting algorithms are based on Markov models where observations of topics are trained and Topics with a similar life cycle are recorded and they share the same model such as Zeng et al. 2007; Liu and Guo 2011. Budak et al. (2011) proposed two trends definitions called coordinated and uncoordinated trends that detected topics that are highly popular among the clustered and distributed users, respectively.

1. CHALLENGES OF SOCIAL MEDIA ANALYTICS

As SMA is an emerging research field, it faces a number of challenges. Similar to the social network analytics, it has theories from network science, sociology, statistics, and graph theory but it lacks a theoretical core and the level of interdisciplinary research co-operation still tends to be low. From a methodological perspective, SMA has a number of challenges like social media data, a collection of data and existing analytics and mining methods.

First, data is generated in large quantities and is highly dynamic and complex in nature. So processing the data with traditional methods or database management's tools is very difficult. In addition to it can include structured and unstructured data characteristics. Second, Boyd and Crawford (2012) have pointed out the issues related to unreliable large data sets from the internet which may lead to inconsistency and incompleteness (when multiple data sets are used together. Regardless of size, data sets are subjected to its limitations and are formerly considered as biased. Third, regarding the data collection, it is a major challenge to gather a large amount of data at one place from a lot of social media platforms as some platforms don't provide the medium to access the data. In addition, the fast-changing nature of social media platforms and the interventions by the users (like twitter has tags) is another biggest challenge. Finally, privacy issues are present when data are to be collected. Researchers and other parties gathering the data may face problems like it is authorized to gather the data and processor to report the data on social media if they are actually public in nature. Regarding the data analysis and mining, there is a lack of systematic approach and frameworks as some of the methods are still in the early stage of development.

In particular, the major weakness lies in the existing computational methods and algorithms.

1. FUTURE RESEARCH DIRECTIONS AND INTERDISCIPLINARY CO-OPERATION

To tackle the research challenges mentioned above, there are some future research directions that can be implemented on an interdisciplinary basis. First, SMA research should be principally concerned with developing and evaluating technical frameworks scientific methods, and software tools and platforms for analyzing, modeling and mining large-scale social media data, principally classification based on machine learning based off social media textual content and acknowledgment of social network patterns need upgrading. Most prominently, SMA research should involve the highly dynamic properties of social media data as well as the rapidly changing nature of the designs and concepts of social media applications into contemplation. In particular, IS research focuses on the design, implementation, use, and management of social media as socio-technical systems as well as on their practical impacts may greatly benefit from SMA.

Based on these methods, frameworks, and toolsets, SMA research can serve as another direction address the development of corresponding data driven and dynamic decision making or decision aiding frameworks, particularly in the business context like SMA research can help extend BI by adding a "social" component, which can be referred to as "social BI". More specifically, as business-related decision-making frameworks require well communicative and noticeably defined performance measures, SMA research should elaborate on the issue of measurement, which is highly relevant in IS research. In anticipation of now, quantifying performance measures is very challenging, due to a broad spectrum of different social media applications present among other things. The consequence of this measurement problem is the difficulty for enterprises to determine social media return on investment. There are still release questions on how to measure the effectiveness of public relations and advertising activities through social media, which is relevant for making decisions on the suitable targeting and budget allocation for such activities, as well as in the collection of the instruments.

Dealing with the provision of architectural designs and solution frameworks for existing and new social media applications, based on research findings from the study of the interaction between the design and concept of social media applications and the eventual adoption and usage behavior, as well as from other insights and practical implications derived from analyzing social media data can be another research stream. IS research will benefit when it comes to designing and implementing new social media-based applications and information system.

Finally, SMA should not be limited to analyzing social media only instead, it also should address the interplay between social and traditional media, that should include other non-social media online content and activities such as an online advertising campaign may inspire consumers' production of user-generated content and social media activities and vice versa.

In this contribution, we call for a major increase in the level of interns disciplinary research collaboration. This collaboration should not be partial only to scattered collaborations between individual researchers, but must extend to large-scale, synchronized research actions across all major disciplines that are stakeholders in social media research, including sociology, information systems, linguistics, media and communication studies, trade studies, economics, political science, and social psychology, computer science, and statistics. This attempt must aim to create momentous advancements in the scientific methods for analyzing social media, as well as to answer all research questions from across the different disciplines like in communication studies, theories on the subject of public specialty and media possessions such as agenda disintegration, setting, in sequence circulation, judgment leadership, the corkscrew of silence, reflection, and split hypotheses, the digital divide, and inequalities in the distribution of attention and influence should be adapted to the special conditions of the Internet and tested. More specifically, theories from the social sciences should be connected to the methods of the applied sciences such as computer science, information systems, and statistics, and vice versa.

CONCLUSIONS

In this paper, SMA is introduced as an emerging interdisciplinary research field that, in my view, will significantly influence future social media-related research from different disciplines, and will have a very high practical relevance. Furthermore, I believe that SMA can help IS research to develop decision-making or decision-aiding frameworks for tackling the issue of social media-related performance measurement as well as to provide architectural designs and solution frameworks for new social media-based applications and information systems. Finally, we call for an interdisciplinary SMA research agenda as well as a significant increase in the level of inter-disciplinary research co-operation, which must aim to generate significant advancements in scientific methods for analyzing social media, as well as to answer research questions from different disciplines.

ACKNOWLEDGEMENT

REFERENCES

1. Boyd d, Crawford K (2012) Critical questions for big data. Information, Communication & Society 15(5):662–679
2. Bruns A, Burgess J (2012) Local and global responses to disaster: #eqnz and the Christchurch earthquake. In: Proc of Australia New Zealand disaster and emergency management conference AST management, Brisbane, pp 86–103
3. Bruns A, Highfield T (2013) Political networks on twitter: tweeting the Queensland state election. Information, Communication & Society 16(5):667–691
4. Bruns A, Liang EY (2012) Tools and methods for capturing twitter data during natural disasters. First Monday 17(4

5. Bruns A, Stieglitz S (2013) towards more systematic twitter analysis: metrics for tweeting activities. International Journal of Social Research Methodology 16(2):91–108

6. Budak C, Agrawal D, El Abbadi A (2011) Structural trend analysis for online social networks. Technical report UCSB/CS-2011-04,UCSB

7. Burgess J, Bruns A (2012) Twitter archives and the challenges of "Big social data" for media and communication research. M/C Journal 15(5):1–7

8. Facebook (2013) Facebook key facts. http:// newsroom.fb.com/Key-Facts. Accessed 2013-02-19

9. Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proceedings of the National Academy of Sciences USA 99:8271–8276

10. Global Web Index (2013) Twitter now the fastest growing social platform in the world.
    https://www.globalwebindex.net/ twitter-now-the-fastest-growing-social-platform-in-the-world/. Accessed 2013-09-03

11. Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. MIS Quarterly 28(1):75–105

12. Kasiviswanathan SP, Melville P, Banerjee A, Sindhwani V (2011) Emerging topic detection using dictionary learning. In: Proc of CIKM'11

13. Kavanaugh A, Fox EA, Sheetz S, Yang S, Li LT, Whalen T, Shoemaker D, Natsev P, Xie L (2011) Social media use by the government: from the routine to the critical. In: Proc of 12th annual international digital government research conference, College Park, Maryland, pp 121–130

14. Krippendorff K (2004) Content analysis: an introduction to its methodology. Sage, Lon-don

15. Larson K, Watson RT (2011) the value of social media: toward measuring social media strategies. In: Proc of 32nd international conference on information systems, Shanghai, China

16. Liu B (2011) Web data mining: exploring hyperlinks, contents, and usage data. Springer, Heidelberg