Framework for Participative and Collaborative Governance using Social Media Mining Techniques

Nazura Javed Research Scholar Bangalore University Bangalore, India

ABSTRACT

Social media and networking sites have broken the barriers in communication and brought about a revolution in information access, dissemination and communication. Given the inclusive and collaborative nature of social media, political leaders and government organizations too have attempted to harness it for building support, gauging popularity and analyzing the opinions of citizens. This paper, proposes a framework which focuses on the application of text mining techniques on social media content for achieving participative, collaborative and inclusive governance. Most of the political entities and government departments have twitter accounts. Though the tweet text is limited to 140 characters, the accompanying metadata and hashtag (#tag) embedded content make it a potent source for extracting entities, concepts and topics. Web Crawling for content pertaining to these entities and compilation of the same, results in a corpus suitable for text mining. Applying Natural Language Processing (NLP) and mining techniques like Part-of-Speech (POS) tagging, classification, clustering to this generated corpus enables categorizing and summarizing: a) sentiments b) queries and c) grievances. Summarized view of the content makes it amenable for decision making and formulation of responses. A Knowledgebase (KB) is created and perpetually updated with the concepts, entities, summaries, queries and responses. The continued process of creation and enhancement contributes to machine learning. The scope of this paper is limited to proposing a conceptual framework for social media mining facilitating informed decision making for better governance.

General Terms

Social Media, Analytics, Text Mining, Governance

Keywords

Framework, Social media, Text Mining, Political, Machine learning

1. INTRODUCTION

Social Media has broken down barriers in communication. Social Networking Sites (SNS) have provided a platform for the user community to interact, collaborate, promote and debate. They have emerged as forum to solve the issues and debate solutions. Social media mining and analytics have assumed relevance today because of the increasing volume of e-content and the imperative need for gaining insight into Muralidhara B.L. Bangalore University Bangalore, India

diffusion of information, opinions, sentiments and trends. The unstructured, voluminous and multiplicative nature of content however, makes analytics an exigent task. Thus suitable techniques and tools need to be devised for processing and analytics.

Social media is increasingly used in administration and governance. Micro-blogging services and Social Networking Sites (SNS) have increased citizen participation. Responsiveness, transparency, collaboration and mass participation are essential characteristics of collaborative governance. These characteristics are inherent in social media. The governments of nations, public sector companies and research organizations have recognized the need to channelize it to improve services and communication with citizens. The literature survey undertaken by us indicates that most of the research in this area is confined to: a) usage of social media for campaigning, image building and garnering support b) sentiment mining with a view of gaining insight into the sentiments of the citizens with respect to political parties and leaders and c) effectiveness of social media as a tool for information dissemination. Hence through this paper, the possibility of establishing an interactive two-way communication between the government and the citizens is explored.

This paper proposes a conceptual framework to summarize public opinions, derive Frequently Asked Questions (FAQs) and compile concerns of the people by using concept extraction techniques, text analytics and mining. It focuses on the following aspects:

- a) Twitter as a primary source to assimilate popular topics and entities.
- b) Looking beyond the scope of sentiment polarity determination and proposing sentiment summarization and segregation.
- c) Clustering the queries in order to derive FAQs.

Fig 1 provides an overview of the social media mining and collaborative networks.



Summarized View

Fig 1: Overview of Social Media Mining and Collaboration

The steps can be stated as follows:

- a) The government disseminates information about its policies, schemes, drives, laws, rules and regulation through twitter, blogs and social networking sites. Government encourages crowd sourcing and participation.
- b) The citizens (user) participate, raise queries, discuss and debate on social media.
- c) The user content is compiled into corpus suitable for text mining algorithms.
- d) Content is classified into sentiments, queries, grievances or concerns using NLP techniques.
- e) Content is further grouped and summarized using clustering techniques.
- f) Government views categorized, summarized content and queries. Dash board style view provides an overview of the citizen opinion and queries.
- g) The social media sites are updated with the responses for FAQs.
- h) The knowledgebase is updated, enhanced with categorized and summarized content, queries and corresponding responses.

The remainder of the paper is structured as follows: In section 2 we overview the related literature. Section 3 proposes the methodology for mining entities, concepts, sentiments, opinions and queries from the social networking sites. Section 4 defines the boundaries, scope and concludes by examining the future scope for enhancement and research.

2. RELATED WORK

The advent of Web 2.0 has brought about a revolution in the field of Information and Communication technology and participative governance. The relevance of social media and their role in e-governance are highlighted in the following literary works. The role of the social networking sites like Facebook, Google+, micro-blogging services such as Twitter, wikis, and media sharing sites such as YouTube and Flickr in promoting interaction, collaboration, permeation and dissemination [1]. Information and Communication Technology (ICT) is cost-effective and convenient means to promote openness, transparency and to reduce corruption. The public sector too has recognized the role of the social media for responsive, collaborative and transparent governance [2, 3, 4]. The role of ICT in channelizing, mobilizing the nation's resources and empowering its human resource cannot be disputed. ICT and social media promote transparency, responsiveness and participation. They can be used to help citizens to become co-creators and co-producers of the services they use [5].

The sheer volume of content on the social network has necessitated the use of automated techniques to decipher the information and extract relevant knowledge. Text analytics and mining techniques are used for extracting knowledge from unstructured and semi structured content. Supervised and Unsupervised learning techniques are used for machine learning and information retrieval. Since the unlabeled data are easier to obtain, techniques that make use of the positive and unlabeled data for learning are examined [6]. Text mining techniques for mining sentiments and opinions from social media are used [7, 12].

Discovering the connections between words and emotions. can help to improve emotion prediction. Modeling these connections can make emotion prediction efficient [9]. Opinion mining, a sub-discipline within data mining refers to the computational techniques for extracting, classifying, understanding and assessing the opinions expressed on various online news sources, social media comments and other generated content. Sentiment analysis is often used in opinion mining to identify sentiment, effect, subjectivity and other emotional states in online text. MuSES, a multilingual sentiment identification system, leverages knowledgebase like Wikipedia, and labels word-level sentiment for non-English words. A scoring-based sentiment algorithm that assigns numeric scores to phrases and a novel rule-based algorithm that considers emoticons are developed [8]. The use of text mining techniques to investigate hidden pattern consumers' attitudes towards global brands is also studied in [10]. The study found that Twitter can be used as a reliable method in analyzing attitudes towards global brands. Algorithms for extracting heterogeneous comments that represent the different aspects and sentiments in the news articles are proposed [11]. Mining sentiments from natural language is an

extremely difficult task. The Senti-Tut project examines the use of irony and its role in polarity reversal [13]. The organizations can stay connected to guests and address their issues through social media. Mining techniques can be used for extracting and categorizing feedbacks from the customers. Social media can be exploited by organizations for branding, sales, product development and innovation [14]. Mining social media poses challenges because of the use of abbreviations, phonetic substitutions and structure. Thus it is necessary to normalize and convert the content into standard format [15].

Knowledgebase is at the heart of any powerful machine learning system. Techniques are developed for building a knowledgebase that captures the entities and relationships from social media [16]. Semantically similar information can be expressed using different linguistics. Learning the linguistic variability in text corpora for the purpose of automatic annotation is discussed [17].

Micro-blogging sites like Twitter serve as a channel for disseminating political content and engaging political dialogs. Social media represents an ideal vehicle to gauge public opinion on policies and political positions as well as build community support for candidates contesting for public offices. Social media is an ideal platform to spread political opinion publicly. Political institutions have begun to use Facebook and Twitter to enter into direct dialogs with the citizens. Wall posts and corresponding comments on Facebook let people express and discuss their opinions and sentiments on a given news, topic or person. Usage of social media for campaigning, image building and garnering support is explored [19, 23]. A framework for data tracking, monitoring and data analysis for micro-blogging, SNS and weblogs in political context is proposed [20]. Political parties can be identified through their characteristics. The political leanings of the users of social media can also be discovered [21]. Internet is affecting how people engage with each other and with the larger institutions of society. Internet represents a new means of building social capital. Who uses the Internet, and how they use the Internet, is significant in predicting political participation [22]. Apart from using social media for communicating and disseminating information to the citizens, the phenomenon of crowd-sourcing has gained relevance. The governments make use of specialist as well as local knowledge of citizens in order to improve their administrative processes and public services [24]. The potential of drawing in citizens through online consultations, crowd-sourcing and text mining need exploration. Platform for publishing policy decisions and synthesis of opinion regarding the policies for the purpose of decision making are developed [25].

This paper proposes a framework for: a) mining and discovering popular topics from twitter on the basis of the content and metadata, b) synthesizing and summarizing public opinions and (c) formulation of FAQs and segregation of feedbacks into sentiments and grievances.

3. METHODOLOGY

Social Media can bring about inclusive governance by facilitating citizen involvement and participation. The

presence of government entities on social media can make a significant impact. Social Media is mainly used by the governments for the following:

- a) Campaigns and initiatives for the betterment of society.
- b) Broadcasting information about policies, rules and regulations.
- c) Dissemination of information for calamity and disaster management.
- d) Making citizens aware of the public welfare programs and schemes.
- e) Polling, inviting opinions and crowd sourcing.
- f) Providing linkages to national portals and government websites.

The objective of the proposed framework is to extract, categorize and summarize the social media content. The concise, categorized, summarized content provides valuable insight into the social scenario. Visualization and summarization facilitates the generation of responses. This proposed framework would comprise of the following 3 processing components:

- Generating Twitter Corpora: Twitter has assumed significance in the political arena with most of the political and government entities leveraging it for promoting themselves and their initiatives. Twitter is a micro-blogging service that allows people to communicate with short 140 character messages. Tweets can be collected using the RESTful or Streaming API. The accompanying metadata contains valuable statistics like the number of retweets, date and time of creation, location, followers count etc. Statistical Analysis of the metadata brings to fore the most popular topics and entities. The #tag and @tags in the tweet facilitate the extraction and segregation of topics and entities from the tweets. The extracted topics function as keywords for web crawling and searching the related web content. Compiling topics and content yields corpus suitable for text mining.
- b) Text Analytics and Mining: NLP techniques like POS tagging can be used for chunking and annotating the text. The lexicons like WordNet facilitate the grouping of semantically similar content into preliminary classes. Clustering in conjunction with sentiment lexicons like SentiWordNet makes it possible to derive clusters of sentiments. Each cluster can be further assigned a relevant class label to reflect the sentiment class. Summarization of queries yields FAQs.
- c) Presentation techniques like dash boards provide lucid picture of the political scenario. Analysis, decision making and response generation are facilitated.

Fig. 2 depicts the essence of the proposed framework.

Stage I	Stage II	Stage III	Stage IV	Stage V
Corpus of tweets	KB: Topics, entities	KB: Concept + Text	KB: Classified content	Classified, summarized view
I Entity Extraction Extract @ and # tagged text	II Information Retrieval	III Semantic Classification Text Pre-processing	IV Clustering	V Response Generation View Summaries
Perform statistical analysis of metadata for determining relevance Compilation of topics and entities	Organize and structure content	POS tagging and subject, predicate, query determination Perform semantic classification using lexicon	sentiments using sentiment lexicons like SentiWordNet Cluster the queries to derive FAQs Refine class labels to reflect clusters	and FAQs Obtain/ generate responses for FAQs Update/Enhance knowledgebase
KB: Topics, entities	KB: Concept + Text	KB: Classified content	Classified, summarized view	KB: Updated, Enhanced

Fig 2: The framework for social media mining for collaborative governance

The framework for mining social media for participative and collaborative governance comprises of 5 major processing stages.

Stage I. Extract Entities, Concept(s), Topic(s)

- a) Input: A corpus of tweets and metadata.
- b) Process:
 - i. Extract # tagged and @ tagged content from the tweets and identify entities, topics or concepts.
 - ii. Perform NLP to extract the noun phrases and establish the relevant entities.
 - iii. Perform statistical analysis on the accompanying metadata to determine the relevance of entities.
- c) Output: Knowledgebase with keywords representing entities, topics or concepts.
- Stage II. Information Retrieval
- a) Input: The knowledgebase comprising of entities and topics from stage I.
- b) Process:
 - i. Use web crawling to retrieve web content for the input entities and topics.
 - ii. Compile topics and content to build a corpus.
- c) Output: An organized corpus of topics, entities and textual content.

Stage III Semantic Classification

- a) Input: Corpus from Stage II.
- b) Process:
 - i. Pre-process the corpus so as to make it suitable for Text Mining.

- Use POS tagging techniques to determine a) parts of speech b) subject and predicate and c) queries.
- iii. Use lexicons like WordNet to arrive at synonyms.
- iv. Use customized distance-based classification algorithms to perform semantic classification of the Corpus.
- c) Output: Corpus with Class Labels indicating the categories of content.

Stage IV Clustering

- a) Input: Classified content from Stage III.
- b) Process:
 - i. Cluster the classified content in conjunction with sentiment lexicons like SentiWordNet such that each cluster represents similar sentiments and opinions.
 - ii. Cluster the queries to derive FAQs.
- c) Output: Redefined classes which reflect sentiments, queries and grievances.

Stage V Response Generation

a) Input: Classified Content from Stage IV.

- b) Process:
 - i. Organize the display of the summaries and FAQs.
 - ii. Accept or generate responses for FAQs.
- c) Output: Updated Knowledgebase

4. CONCLUSION AND FUTURE WORK

In this paper, framework to mine the social media content in political context is proposed. The investigation dealt with the following: a) leveraging Twitter as a source of mining entities, concepts and topics and b) framework for text mining with a purpose of segregating and summarizing sentiments, queries and grievances. This work assumes relevance as it analyzes the societal scenario and facilitates decision making and generation of responses. Multilingual support is however not explored in this paper. The entities, topics or concepts are limited to those appearing in the tweets. Though emoticons and their interpretations will be considered for future implementation, the rules to identify and decipher irony in the posts are beyond the scope. Generation of responses may require human intervention, but as the system learns and the knowledgebase is enhanced, the autonomy of the system improves. The applicability of this framework is not just limited to participative and collaborative governance. It can be applied by corporate business houses to analyze public response, the popularity of their products or services and respond to the queries. Implementation of the above conceptual framework for analytics and response generation is the next milestone.

5. REFERENCES

- [1] Magro, M. J. 2012. A review of social media use in egovernment.Administrative Sciences, 2(2), 148-161.
- [2] Bertot, J. C., Jaeger, P. T., & Grimes, J. M. 2010. Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. Government information quarterly, 27(3), 264-271.
- [3] Unsworth, K., & Townes, A. 2012. Social media and E-Government: A case study assessing Twitter use in the implementation of the open government directive. Proceedings of the American Society for Information Science and Technology, 49(1), 1-3.
- [4] Arpit, B. 2012. E-Government and Social Media as Openness and Anti-Corruption Strategy. Research Journal of Management Sciences, 1(1), 48-52.
- [5] Megele C 2012. How to move from local e-government to collaborative e-governance, the guardian, http://www.theguardian.com/local-government-network/ 2012/may/08/local-government-e-citizenship. Accessed 12 December 2013
- [6] Zhou, K., Gui-Rong, X., Yang, Q., & Yu, Y. 2010. Learning with positive and unlabeled examples using topic-sensitive PLSA. Knowledge and Data Engineering, IEEE Transactions on, 22(1), 46-58.
- [7] Kavanaugh, A. L., Fox, E. A., Sheetz, S. D., Yang, S., Li, L. T., Shoemaker, D. J., ... & Xie, L. 2012. Social media use by government: From the routine to the critical. Government Information Quarterly, 29(4), 480-491.
- [8] Xie, Y., Chen, Z., Zhang, K., Cheng, Y., Honbo, D. K., Agrawal, A., & Choudhary, A. N. 2014. MuSES: Multilingual Sentiment Elicitation System for Social Media Data. Intelligent Systems, IEEE, 29(4), 34-42.

- [9] Bao, S., Xu, S., Zhang, L., Yan, R., Su, Z., Han, D., & Yu, Y. 2012. Mining social emotions from affective text. Knowledge and Data Engineering, IEEE Transactions on, 24(9), 1658-1670.
- [10] Mostafa, M. M. 2013. More than words: Social networks' text mining for consumer brand sentiments. Expert Systems with Applications, 40(10), 4241-4251.
- [11] Giannopoulos, G., Koniaris, M., Weber, I., Jaimes, A., & Sellis, T. 2014. Algorithms and criteria for diversification of news article comments. Journal of Intelligent Information Systems, 44(1), 1-47.
- [12] Lloret, E., Balahur, A., Gómez, J. M., Montoyo, A., & Palomar, M. 2012. Towards a unified framework for opinion retrieval, mining and summarization. Journal of Intelligent Information Systems, 39(3), 711-747.
- [13] Bosco, C., Patti, V., & Bolioli, A. 2013. Developing corpora for sentiment analysis: The case of irony and senti-tut. IEEE Intelligent Systems, (2), 55-63.
- [14] He, W., Zha, S., & Li, L. 2013. Social media competitive analysis and text mining: A case study in the pizza industry. International Journal of Information Management, 33(3), 464-472.
- [15] Han, Bo, and Timothy Baldwin. 2011. Lexical normalisation of short text messages: Makn sens a# twitter Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, 2011.
- [16] Chai, X., Deshpande, O., Garera, N., Gattani, A., Lam, W., Lamba, D. S., ... & Doan, A. 2013. Social Media Analytics: The Kosmix Story. IEEE Data Eng. Bull., 36(3), 4-12.
- [17] Boella, G., Di Caro, L., Ruggeri, A., & Robaldo, L. 2014. Learning from syntax generalizations for automatic semantic annotation. Journal of Intelligent Information Systems, 43(2), 231-246.
- [18] Nguyen, L. T., Wu, P., Chan, W., Peng, W., & Zhang, Y. 2012, August. Predicting collective sentiment dynamics from time-series socia media. InProceedings of the first international workshop on issues of sentiment discovery and opinion mining (p. 6). ACM.
- [19] O'Hara, K. 2013. Social machine politics are here to stay. Internet Computing, IEEE, 17(2), 87-90.
- [20] Stieglitz, S., & Dang-Xuan, L. 2013. Social media and political communication: a social media analytics framework. Social Network Analysis and Mining, 3(4), 1277-1291.
- [21] Boutet, A., Kim, H., & Yoneki, E. 2013. What's in Twitter, I know what parties are popular and who you are supporting now!. Social Network Analysis and Mining, 3(4), 1379-1391.

International Journal of Computer Applications (0975 – 8887) Volume 123 – No.8, August 2015

- [22] Gainous, J., Marlowe, A. D., & Wagner, K. M. 2013. Traditional cleavages or a new world: Does online social networking bridge the political participation divide?. International Journal of Politics, Culture, and Society, 26(2), 145-158.
- [23] Vergeer, M., & Hermans, L. 2013. Campaigning on Twitter: microblogging and online social networking as campaign tools in the 2010 general elections in the Netherlands. Journal of Computer-Mediated Communication, 18(4), 399-419.
- [24] Grubmüller, V., Götsch, K., & Krieger, B. 2013. Social media analytics for future oriented policy making. European Journal of Futures Research, 1(1), 1-9.
- [25] Charalabidis, Y., Loukis, E., Spiliotopoulou, L., & Diamantopoulou, V. 2013. A Framework for Utilizing Web 2.0 Social Media for Participative Governance. InEuropean, Mediterranean & Middle Eastern Conference On Information System 2013, 01 (Vol. 4).
- [26] Berry M.W., Castellanos M. 2007 Survey of Text Mining: Clustering, Classification and Retrieval Springer, Second Edition September 30, 2007
- [27] Russell M. A. 2013 Mining the Social Web O'Reilly, ISBN:978-1-449-36761-9