Review on Comment Volume Prediction

Mandeep Kaur[1,] Prince Verma[2]

Computer Science engg.CTIEM Jalandhar,Punjab

Abstract: *In this paper we present the concept of social media and its various functional building blocks. Social media has become an ubiquitous part of social networking and content sharing. Social media make use of mobile and web to create the platforms with which individuals and communities share, discuss, and modify user-generated content. This is examined that social media user's comments are also very much influential in order to make or change people's perception and to bring some topic in trending. Different techniques have been proposed for comment volume prediction which are discussed in the literature survey.*
Keywords: *Social media, Conversations, Sharing, Reputation, Comment volume.*

I. INTRODUCTION

Initially Internet was explored for limited purposes as for reading, watching or shopping only. But today's consumers are more intelligent and are utilizing platforms such as blogs, content sharing sites, wikipedia and social networking to create, converse and share Internet content. SixDegrees.com was the first social network site, launched in 1997. It allowed the users to create their profiles, list their contacts and surf the Friends lists only.

But now, there exists a rich ecology of social media sites, that vary in terms of their functionality and scope. Some of them are for the general purposes, like Facebook, Hi5, and Friendster, where MySpace, YouTube are media sharing sites which focus on shared videos and photos. While other sites such as LinkedIn are meant for professional networks. Even weblogs after a slow progress in the late 1990s have become very popular, as they are easy to create and manage. The authors of weblogs range from a normal person to professional writers and celebrities. Now a days, the 'blogosphere' has become a significant source of public opinion with more than 100 million blogs and their interconnections. Even there are search engines, i.e. Technorati, that are dedicated to searching blogs[20]. Twitter is also getting popular as there more than 300 million active users are available on twitter by 2016.

Social networking sites have become a popular platform for people to share information and communicate with others, which can significantly affect a firm's reputation, survival and sale's revenues. These services are acting as a multi-tool with routine applications e.g. news, advertisements, communication, commenting, banking, marketing etc.

Facebook is getting the position of being most popular social networking site, as in Facebook, the daily data ingested into the databases, is beyond 500 terabytes, In one of the FB's largest Hadoop clusters there is 100 plus petabytes of disk space and the content items shared per day is about 2.5 billion [1].
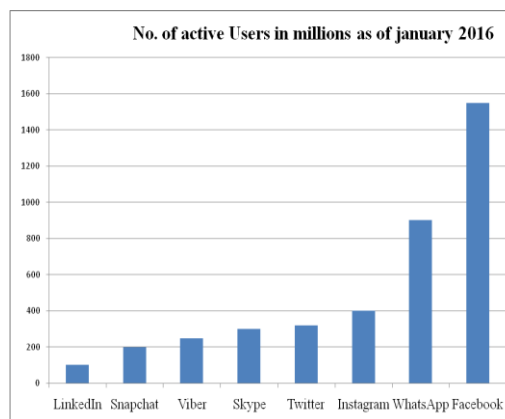

Fig. 1. Number of active users in millions and social networks

Figure 1. gives the information on the most popular networks worldwide as of January 2016, ranked by the number of active accounts. Active users are those which have logged in during the last 30 days. With a constant presence in the lives of their users, social networks have a strong social impact.

II. SOCIAL MEDIA FUNCTIONAL BLOCKS

Different levels of social media functionalities are mainly constructed from these basic seven blocks:

Fig. 2. Social media functionality constructs

Identity: It is the extent to which users reveal themselves to others. The information that user disclosed can be name, gender, age, location, profession or the information that portrays him in certain ways. There are many different social media platforms which are built around identity and require users to set up profiles like Facebook. As identity is the core to different social media sites, the major issue that needs attention is privacy. Though users willingly make public their identities on various social media sites, but they also have serious concerns about how other firms view and use their information for data mining and the extent to which social media platforms passively or actively encourage such activities. This has resulted in initiating various class-action lawsuits by different users and government agencies, in order to keep privacy of a user.

Conversations: It is the extent to which users converse or communicate with each other. The prime purpose of many social media sites is to facilitate conversations among individuals and groups. Different kinds of conversations can happen for different reasons as people tweet, blog or update status to remain in touch of new ideas or current topics or to meet new like-minded people. In this way, people see social media as a platform of making their message heard and impacting environmental problems, humanitarian causes, economic issues, or political debates. For instance, Twitter is focused around exchanging small messages which are mostly the updates of real-time statuses, to create an 'ambient awareness' of issues. On the other side, Blogs offer lengthy conversations which can be traced back by staying connected asynchronously. The firms view the 'conversation velocity' by examining the rate and the direction of change in a conversation. By the rate of change of conversation is the count of new conversations over a specific period of time, while the direction of change refers to the continuity or discontinuity of the conversation.

1) *Sharing:* It is the extent to which users send, receive or exchange data. Sharing provides a way for interaction on social media platforms, but it depends on the functional objectives of the social media platforms, whether sharing leads users further to converse or build relationships with each other. Different sites have different objects of sociality as Indie music for MySpace, pictures for Flickr and careers for LinkedIn. For instance, YouTube as a social media platform allow the users to upload and share videos, so object medium is video. But this requires users to get registered themselves and agree the conditional terms in order to upload a video which includes a content management system that makes the content owners able to identify and manage their content, even there are number of employs who are engaged in monitoring and removing content from You Tube in case of violation of the terms.

2) *Presence:* It is the extent to which users know about the existence of others as if others are available or accessible. Presence-focused platforms center on geographical spaces as Friends Around Me facility allow users to share their updates and check-ins of their current location, across the networks i.e. Facebook, Foursquare, Twitter, and displays where friends are and which of them are nearby. But this function is not important for all social media platforms as for instance, in LinkedIn, knowing which users are online or where they are located does not make any difference.

3) *Relationships:* It is the extent to which users are related to each other i.e. form of association among users that leads them to share objects, converse or simply to be in each other's friend list. There are some relationships which are formal, regulated, and in structured way as in LinkedIn allows users to find how they

are linked to each other and how many degrees they are away from a target user, possibly an employer who would like to meet an employee. On the other hand, there are platforms in which relationships are informal and without any specific structure as in Blogs, which allow users to develop a relationship, without any formal rules of how much information they should share with each other. In other cases like Twitter and YouTube in which relationships hardly matter.

4) **Reputation:** It is the extent to which users know about the social standing of each others. The meaning of Reputation varies with the different social media platforms, for instance, LinkedIn builds the reputation of one user based on endorsements from other users. And on YouTube, the reputation is based on count of views, of videos while on Facebook, this depends on 'number of likes' but on Twitter it depends on the 'number of followers'.

5) **Groups:** It is the extent to which users can build communities and upto which levels these are formed. More social a network, the bigger the group of friends and followers is. There are two types of groups mainly i.e. first, 'self-created groups' in which individuals sort their contacts and place their followers, friends or fans into the groups like lists in Twitter. Secondly, groups with administrators who create/manage, invite and approve applicants in the group, which can be open to anyone, closed i.e. user need to request to join group, or secret group in which only invited members are allowed as in Facebook and Flickr.

III. SOCIAL MEDIA COMMENTS

Comments are the opinions of social media users about a document. These are usually short textual messages referring to the main text of the document or to other feedbacks.

A. *Importance of User's comments*

The nature and valence of user's comments on various social media platforms plays a significant role in building or changing the one's perceptions regarding some specific topic or to making it popular. Even, in today's era, it has an important place in various fields whether it is education, sales or predictions etc.

- Perceptions--The paper[15] examined that the perception of people is influenced by the comments of users on social media. The findings were explored in the context of relationship status, in which a person made an announcement about the formation or breakup of a real-world relationships on Facebook. The results suggested that the comments from different users affects the perceptions of a Facebook relationship status update. It showed that the positive comments lead to favourable attitudes whereas negative comments lead to poorer attitudes towards the status. And also the observer's attitude towards a relationship status are more driven by the count of the comments than the real nature of the status, as positive facebook status can be seen as negative if the associated comments are negative in nature.

- Education-This work[16] investigated the effects of using a social media web application to assist learning and teaching in classrooms on learners' performance. By examining the ratio of number of Facebook posts and comments for educational purposes to the one for non-educational purposes, a conclusion is drawn that the use of Facebook does have an impact on student's learning performances, as the students who spent more time on education-related posting and commenting earned better grades.

- Predicting the future-This paper[17] demonstrate that the contents of social media can be used for the predictions of real-world outcomes. Predictions were performed on platform Twitter.com, using almost 3 million tweets. It predicted the box-office revenues of movies by constructing a linear regression model, in advance of their release dates. The results outperformed in terms of accuracy and concluded that a topic's rank in future is strongly correlated with the amount of attention it receives on social media.

- Driving force in s/w evolution- User feedback is important in improving software Quality. Many software companies collect data on user satisfaction through various means. The user's feedback available on the sites thereby supply software companies with a rich source of information that can be used to improve future releases. Even Apple's App Store stated that the newest release addresses many of the issues raised by users. Therefore, user feedback can be and has been the driving force in software evolution [13].

- Trending- It shows a list of topics that have recently spiked in popularity which is evaluated based on the volume of the response towards a particular topic/issue or product.

- Sales and Marketing- McDonald's Australia is serving over 1.7 million people across more than 940 Australian restaurants each day. This global fast food chain has used Facebook video ads to share its unique Australian story, connecting with more than 5 million people in just 5 weeks[18]. And even Coca-Cola Korea has used age group targeting and video ads to connect with 4.5 million young consumers, in order to promote a new pineapple beverage, Sunny 10, which increased its purchase intent for the product by 2 points.

IV. COMMENT VOLUME PREDICTION

Comment Volume on social media can be measured as the count of words in the comment section, the number of comments, the number of distinct users who leave comments, or a variety of other ways. These measures can be affected by various factors like main text of document/post, link to other documents/posts, the time of day the post appears, a side conversation, Page likes, page check-ins, page talking about or page category etc.

For Comment Volume Prediction(CVP), the user comment patterns are modeled over the posts/documents appeared in the past and based on it, predictions are made on the number of comments that the posts/documents are expected to receive in next coming hours. Number of researches are being done in comment volume prediction field, using different social media platforms.

Different Regressor Models as Reduced Error Pruning tree, M5P tree, Multi Layer Perceptron and RBF network can be used to predict the user's comment volume. As paper [1] has worked on facebook, using neural networks and decision trees and developed a software prototype demonstrating the comment volume prediction. It has made evaluations using various dataset variants and concluded that the Decision trees performed well than the Neural Networks in comment volume predictions. Similarly, paper [2] has developed an industrial proof-of-concept demonstrating the automatic analysis of the documents on Hungarian Blogs. In this paper, author has trained various regressor models by considering various features of the blogs and evaluated the results using Hits@10 and AUC@10 measures. The result shows that the regression models outperform than naive models.

Even, classifiers can be used to categorize the comment volumes in specific classes like paper [3].It reports on predicting the comment volume of news articles before their publication using Random Forest Classifier based on the set of five features i.e. surface, cumulative, textual, semantic, and real-world features .It addresses the task in two steps- first binary classification of articles with the potential to receive comments and second to classify articles with 'low volume' and 'high volume'. Outcomes show better results for binary classification and evaluated that the Textual and semantic features are strong performers among others. In the similar way, paper [7] has made analysis on the content and publish time of online news agencies, to detect effective factors of diffusing contents in public. It has also used the Random Forest Classifier to classify articles in three categories i.e. without commented, moderately commented(1-6) and highly commented(>6).The proposed model has made predictions with more than 70% accuracy and reports that the publish date and a weight introduced for content measure, were most informative features. The results can be refined by considering important days (i.e. elections, festivals, holidays) and geographical features in prediction. While paper [5] has shown the dynamics of user generated comments on seven different news websites, using the log-normal and the negative binomial distributions and predicted the comment volume using Linear model and enable comparison across various news sites. The results showed that prediction of long term comment volume is possible with small error after 10 source-hours observations.

The paper [4] has worked on Social bookmarking website Digg.com. By using comment information, it defined a co-participation network between users and studied the behavioral characteristics of users. It measured the entropy and inferred that the users at Digg are interested in wide range of topics. Using a classification and regression framework, it has predicted the popularity of online content based on comment data and social network derived features. It reported a one to four percent loss in classification accuracy while predicting the popularity metric by using only first few hours of comment data as compare to all the available comment data. The results can further be improved by analyzing the polarity of the comments.

Various topic models can be used to extract the hidden topics in post's content. The paper [6] has worked on political blogs using Latent Variable topic model and analyzed the relationship between the content and comment volume. It has also used Naïve Bayes model for binary prediction task i.e. high volume or low volume and evaluated the prediction under the light of precision, recall and F1 measure. It concluded that the modeling topics can improve recall while predicting high volume posts. Even paper [8] has predicted the formation of user-to-content links in Flickr Groups to predict the chance that a user will comment or like an image updated by another user. It has taken into account both the community effect using Transactional Mixed Membership Stochastic Block (TMMSB) model and content effect using Latent Dirichlet Allocation(LDA) for predicting user-to-content links. The time zone effects can be used in future in order to make results more accurate.

Most popular social networking site, facebook is used in paper[9] and presented a data mining architecture to collect social data. This paper has collected different attributes such as about me, comments, wall post and age from facebook using facebook API key, then mined the knowledge to perform the comparison on age group basis for various usages as human behavior prediction, job responsibility distribution, pattern recognition, decision making and product promoting. It has used the K-nearest neighbor algorithm to classify numeric and textual attributes and range value attributes such as wall count, age count, music count and interest count to classify them into various class levels.

Summarization of the user's comments is even more difficult task as these are usually mixed with different opinions, specifically in case of restaurants where different opinions refer to different dishes but evaluated as an overall score of restaurants. Paper [10] has presented a new approach for comment summarization in the context of restaurants. It used the real-world comments, crawled from Yelp and Dianping, the most popular English and Chinese restaurant review web sites. Using the attributes of the dishes and the user's remarks on the attributes as two independent dimensions in the latent space, it constructed a bilateral topic model which is combined with the opinionated word extraction and clustering-based selection algorithms, it provides a high-quality summary on the restaurants as well as the dishes served by the restaurants. This concept can be further used for wider applications like for various selling goods or services.

V.   Conclusion

In this paper, the importance of social media is discussed and it is shown that how it is getting popular day by day. By analyzing the seven building blocks, one can understand how social media platforms vary in terms of their functionalities, in order to develop a good social media strategy by balancing the various building blocks for their community. Different proposed techniques in the literature have been discussed for user's comment volume prediction in order to get the idea of popularity of a topic before its publication. Many researches are being proposed in social media field, but yet there is a huge content left untouched. So the stage is set for social media research to flourish.

References

[1].   Singh Kamaljot, Kaur Ranjeet and Kumar Dinesh, "Comment Volume Prediction using Neural Networks and Decision Trees", 2015 17th UKSIM-AMSS International Conference on Modelling   and Simulation, IEEE, pp. 15-20, 2015.
[2].   Buza Krisztian, "Feedback Prediction for Blogs", Springer International Publishing on Data Analysis, Machine Learning and Knowledge Discovery,2014, pp. 145-152. doi:10.1007/978-3-319-01595-8 16.
[3].   M.Tsagkias, W. Weerkamp, M. de Rijke, "Predicting the Volume of Comments on Online News Stories",CIKM'09 Proceedings of the 18th ACM conference on Information and knowledge management , pp.1765-1768, 2009.
[4].   Jamali, S. and Rangwala, H., "Digging Digg: Comment Mining, Popularity Prediction, and Social Network Analysis", Web Information Systems and Mining, IEEE International Conference,2009, pp. 32-38. doi: 10.1109/WISM.2009.15.
[5].   M. Tsagkias, W. Weerkamp, M. de Rijke, "News Comments: Exploring, Modeling, and Online Prediction", ECIR'2010 Proceedings of the 32nd European conference on Advances in Information Retrieval, Springer, pp.191-203,2010.
[6].   Yano, Tae, and Noah A. Smith. "What's Worthy of Comment? Content and Comment Volume in Political Blogs" In 4th International AAAI Conference on Weblogs and Social Media, 2010.
[7].   Balali, A. and Rajabi, A. and Ghassemi, S. and Asadpour, M. and Faili, H., "Content diffusion prediction in social networks", Information and Knowledge Technology (IKT), 5th IEEE Conference,2013, pp. 467-471. doi: 10.1109/IKT.2013.6620114.
[8].   Negi, S. and Chaudhury, S., "Predicting User-to-content Links in Flickr Groups", Advances in Social Networks Analysis and Mining (ASONAM), IEEE/ACM International Conference, 2012 pp. 124-131. doi: 10.1109/ASONAM.2012.31.
[9].   Rahman, M.M. , "Intellectual knowledge extraction from online social data", Informatics, Electronics Vision (ICIEV), IEEE International Conference, 2012, pp. 205-210. doi:10.1109/ICIEV.2012.6317392
[10]. [Rong Zhang, Zhenjie Zhang, Xiaofeng He, Aoying Zhou, "Dish Comment Summarization Based on Bilateral Topic Analysis" Data Engineering (ICDE), 31st IEEE International Conference, 2015, pp. 483 – 494. doi: 10.1109/ICDE.2015.7113308
[11]. S. M. Tan, P.N, V. Kumar, Introduction to data mining, Pearson Addison Wesley Boston, 2006.
[12]. Z.C. Khan , Thulani Mashiane, "An analysis of Facebook's Graph Search", Information Security for South Africa(ISSA), IEEE, pp.1-8, 2014.
[13]. Laura V. Galvis Carreno, Kristina Winbladh "Analysis of User Comments: An Approach for Software Requirements Evolution",35th International Conference on software Engineering ,USA, IEEE, 2013, pp. 582-591. doi: 10.1109/ICSE.2013.6606604.
[14]. Yu-Hsiu, Hsueh-Yi Lai "Effects of Facebook Like and Conflicting Aggregate Rating and Customer Comment on Purchase Intentions", Springer International Publishing in Universal Access in Human-Computer Interaction. Access to Today's Technologies, 2015,pp.193-200.doi: 10.1007/978-3-319-20678-3_19.
[15]. Paul W. Ballantine , Yongjia Lin, Ekant Veer, "The influence of user comments on perceptions of Facebook relationship status updates", Computers in Human Behavior, Elsevier 2015, Volume- 49, pp.50–55. doi:10.1016/j.chb.2015.02.055
[16]. N. Leelathakul, K. Chaipah, "Quantitative Effects of using Facebook as a Learning Tool on Students' Performance" 10th International Joint Conference on Computer Science and Software Engineering (JCSSE),IEEE 2013, pp.87 – 92. doi: 10.1109/JCSSE.2013.6567325.
[17]. Sitaram Asur, Bernardo A. Huberman, "Predicting the Future With Social Media" Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference, pp. 492 – 499. doi: 10.1109/WI-IAT.2010.63
[18]. https://www.facebook.com/business/success
[19]. http://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/
[20]. Jan H. Kietzmann , K.Hermkens, Ian P. McCarthy, B. S. Silvestre, "Social media? Get serious! Understanding the functional building blocks of social media" Business Horizons, Elsevier 2011, Volume- 54, Issue 3, pp. 241-251. doi:10.1016/j.bushor.2011.01.005