

A Survey of Various Online Social Media & Applications

Abhay Narayan Singh, Anurag Jain

Abstract— Here in this paper a survey of all the technique that are used for the mining of Online Social Media are discussed and their applications such as Community detection and Message Wall Filtering and Student's learning experience and their opinions. Since Online Social Media seems to be advantageous and its usage by various users over internet may create some harm to the other users. Hence a complete survey of all such technique is analyzed and hence by analyzing their various advantages and limitations a new efficient technique can be implemented in future.

Index Terms— www, wcm, wum, Online Social Network, Twitter, Facebook, Co-authorship.

1 INTRODUCTION

World Wide Web (WWW) is a world with rapid and continuous growth in all aspects. It is a data repository which is massive, huge, diverse, dynamic and unstructured. This repository is used as information repository for the purpose of knowledge reference. The challenges that are faced by web are in the form of large, semi structured web pages and also the information on web is likely to be diverse in meaning, quality of the information extracted and the conclusion of the knowledge is obtained from extracted information [1]. Thus for the appropriate perceptive and analysis the data structure of the Web plays an important role for efficient Information Retrieval.

Web mining can be explained as mechanism that categorizes the web pages and internet users in accordance with the contents of the web page and the behavior of the user adopted in the past on the internet. Web Mining is considered as an application of data mining technique. It is generally used to find and retrieve information from the WWW automatically [2].

retrieval of information from WWW in a structured form and thereby for quick retrieval indexes the information. It is explained as the process of extracting useful information from web documents contents. The documents that generally consists of images, audio, video, text files, structured records etc. are processed. WCM technique is used over web documents and the results pages obtained from a search engine. WCM works on the basis of two approaches agent based approach and database based approach. Agent based approach looks relevant information with the characteristics of a domain in detailed manner whereas the database approach regain the semi structure data from the web.

In Web Usage Mining (WUM) the user's profile is recorded and their behaviors in a log file inside the web. Web usage mining process obtains useful information from the data which is accessed from the web by the user while it surfs. It works by extracting data stored in server access logs, agent logs, client-side cookies, user profiles, meta data etc [3].

The basic strategy behind all these process is to gain and obtain information based on the user's behavior and interests and show them the results. Like when resolving a query user's are displayed with multiple results which are generally in the form of URL's of various web pages. For such processes and operation multiple techniques are adopted using multiple algorithms.

Web Mining directly relates with social networks for learning user's behavior and obtaining knowledge from the data being shared. Defining a social network it is a social structure consisting of individuals or having organizations which are in related to each other by some means. The social network standpoint provides a group of methods which that are capable of analyzing the structure of social entities and the theories which can explain the patterns observed in the structures. The study of these structures makes social network analysis that work for identification of local and global patterns, locating the influential entities and examining the network dynamics.

As the internet contains information of various type and kind thus for exploiting and gaining knowledge from this data there are users who dwell over the internet and continuously use it. The users basically share, disseminate and communicate multiple type of information among them-



Fig: 1 Web Mining Categories

The information that can be retrieved by web mining technique is basically divided in classes which are:

- Web content from the data residing on the web pages and inside the documents contained through the web pages.
- Web activity acquired from server logs and browser activity.
- Web graphs acquired from links between the web pages and other various types of data.

Web Content Mining (WCM) generally works upon the re-

selves and with others. The information generally is in the form of texts, audio, video, images etc. [2]. The users also belong to multiple communities consisting of similar type of users in behavior and influencing each other to share different type of data. The medium for this type of data exchange is generally in the form of mails, messages and social networks etc.

Social network advances to understand social interaction which are needed to be first visualized and then investigated through the properties of the relations between the units and not upon unit properties itself.

In a social network there exists different types of relations which may be in singular or combination form with the network configurations and network analytics. While a social networking service provides a platform for building the social networks or social relations by the users and among the users who share interests, activities, backgrounds, relations etc. social network service helps each user to maintain its profile that contains his or her social links and information about other additional services [4].

Therefore Social networks generally enable the users to create a public profile and maintain a list of users for sharing connections and views and even cross the connections inside the social system. Social network services are web based services facilitating the user to interact over the Internet that may be in the form of e-mail servicing and instant messaging. Social network even allow multiple information and communication tools in the form of mobile connectivity, photo, video, sharing, blogging etc.

Social networks provides community services which are group centered and sometimes are considered as individual centered service. Social networks are generally divided into communities that consist of users having same like, features, dislikes etc. Considering over the main types of social networking services includes those which are categorized in groups or communities of users which are say schoolmates, politicians, celebrities etc. and consisting of recommendation system that is directly linked to the trust of its users. Some of the Social networks that are most popular and widely used are Facebook, Google+, YouTube, LinkedIn, Instagram, Pinterest, Tumblr and Twitter.

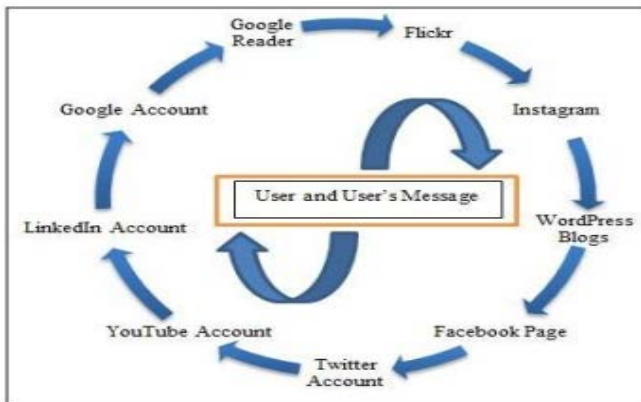


Fig.2 Social Networks

Social network may be acquired by each user may be an online or offline and generally consisting of friends, families and people they are acquainted with. Basic strategy or aim lying

behind of online social networking is to make users social networks noticeable to others who are not in his/her direct network. Social network held its user together by means of kinship, friendship, classmates, colleagues, business partners, etc. which is in the form of pre-established interpersonal relationship. In general the connections in social networks between the users are built one at a time.

The basic reason behind people joining a social networking site is to maintain old relationships with others, find new persons and form new relations for expansion of their network over the web.

Social networking concept is greatly unique in its own way as users can collectively identify others if they are inappropriate, unoriginal or fake and also in social networks users do not compartmentalize their life by having only one social account. Multiple communities in the social network are hold together and sometimes recognized by common interest. The users may possess a hobby for which the community members may be passionate, have a common goal, project, similar lifestyle, geographical location, profession, common interest etc. Thereby in social networks, there are generally two types of users those exhibits and have different influence and different behavior [5].

Thereby Social Network services are web based or internet based group centered services in which users share almost every type of data and can be visualized as an internet service helping the user to build a social network on the internet and relations with other users for purpose of sharing of interests, backgrounds, establish real life connections, participate in multiple activities and discussions with the users that can be characterized or recognized by communities.

A Social network provide base over the internet for maintaining the social associations among the users and helps the users to search other users that may have alike types of interests. It also provides platform for publishing the content and provide knowledge which is provided or generated by other users and also shared, authorized and approved by other users [6].

Social networks enable the present internet generation to maintain interaction with the technology and its usage as well as with other people. OSN's can be well thought-out as a combination of technological, economical and social drives that are capable of fulfilling the need of the users for building social networks, relations etc. over the internet or the web [7].

2 LITERATURE SURVEY

X. Chen [8] et.al. explained that students while being on social media share their experiences and opinions feelings about the learning process. Such type of data is useful for analyzing the student behavior which may directly lead to enhance student learning. With the help of qualitative analysis and large scale data mining techniques students problems can be encountered and analyzed from various tweets and posts. The algorithm can then be trained to detect student's problems. Through this methodology student's experiences can be analyzed. Their methodology basically provides learning analytics, educational data mining, learning technologies. In their research social media data can be analyzed for educational purpose overcom-

ing manual qualitative analysis and large scale computational analysis. With this analysis college students experiences can be encountered which thereby can help increasing the quality of education [8].

Mattias Rost [9] et.al explained that Online services are capable of providing opportunities for understanding human behavior with the help of large aggregate data sets that their operation collects. But the data sets collected by them are not able to model or mirror the world events unproblematically. They have considered the data from Foursquare for analyzing social media as a communicative and not as representational system. They have highlighted four features of Foursquare's use: firstly the relationship between

attendance and check-ins, second event check-ins, third commercial incentives to check-in and humorous check-ins. They proposed that social network data should be viewed and analyzed as communicative data. This means data that is produced as a side effect of communication between users, rather than as a representation of some underlying activity. Looking at the data overall they observed that how the number of check-ins does not have clear correspondence with visitor numbers and the role of promotions as a motivation for check-ins. Also the evidence of humor on the other hand showed itself to be fairly complicated and hardly seemed to affect any large scale analysis [9].

Phil Long [10] et.al. Recognized that the future of education may emphasize new technologies like ubiquitous computing devices, flexible classroom designs, and also innovative visual displays. But big data and analytics is the major factor shaping the future of higher education. Basing decisions on data and evidence is a bit obvious and also research indicates that data driven decision making helps enhancing organizational output and productivity. They explained that Higher education as a field is capable of gathering an astonishing array of data about its "customers," and has conventionally been inefficient in its use of data often operating with substantial delays in analyzing readily evident data and feedback. They proposed that analytics in education should be transformative and the use of analytics can generate privacy, profiling, information sharing, and data stewardship in higher education. The basic proposal by them included that in education the value of analytics and big data can be found in guiding reform activities in higher education, and can assist educators in improving teaching and learning [10].

Roja Bandari [11] et.al. presented that News articles are particularly time sensitive comprising of intense competition among news items for wide propagation. This makes the task of foretelling the popularity of news items on the social web as interesting and challenging. They explained that in previous research prediction of online popularity is based on early popularity. They presented a multi-dimensional feature space consequent from properties of an article and estimate the efficiency of these features that can serve as predictors of online popularity. They observed regression and classification algorithms and demonstrated that even with randomness in human behavior possibility to predict ranges of popularity on social media can have an overall 84% accuracy. They also

studied differences between traditionally prominent sources and those that are immeasurably popular on the social web. They actually predicted the popularity of news items on Twitter with the help of features extracted from the content of news articles. They explained four features that were analyzed - firstly the source of the article, the category, subjectivity in the language and the named entities mentioned. Though in their results these features are not sufficient for predicting the accurate number of tweets that an article will earn but can still be effective for providing a particular range of popularity of the article on Twitter. They have also analyzed that on the basis of number of re-tweets the top news sources on twitter is not conventionally popular news agency. But the most important predictor of popularity is the source of the article. This explains that the readers influenced by the news source that publish the article [11].

Liangjie Hong [12] et.al. reviewed that social network services are now a practicable source of information for the users. Explaining the concept of Twitter they presented that information supposed to be important by the community generally propagates through or via retweets. The characteristics of popular messages can be studied and analyzed for breaking news detection, personalized message recommendation, viral marketing etc. they investigated the problem of prediction of popularity of messages by the number of future retweets and the factors that influence information propagation in Twitter. They put together the specified task into a classification problem and studied features based on the content of the messages, temporal information, metadata of messages and users and the structural properties of the users' social graph on a large scale dataset. With their proposed methodology the prediction of messages is made possible attracting thousands of retweets with good performance [12].

Hila Becker [13] et.al. suggested that social media sites like Twitter and others contain huge amounts of user contributed messages comprising of real world events. Such events messages sometimes contain interesting and useful information like event time, location, participants, opinions that might interest people in learning about an event. They explained that with effective selection technique of quality event helps improving applications like event browsing, search etc. they proposed exploring of representative messages among Twitter messages which corresponds to the same event, which can thereby give high quality, relevant messages capable of providing useful event information. They presented their approach and results on the basis of large-scale dataset of Twitter messages capable of automatic selection of event messages that are relevant and useful. They proposed explanation that sometimes a single event attracts thousands of social media content items, so the proposed methodology should be capable of ranking and filtering event content for applications that aim to communicate that content effectively. They proposed three centrality-based approaches, Centroid as the favored way to select tweets given a cluster of messages relating to an event. Whereas LexRank and Degree tend to select messages that are strongly similar to one another, but sometimes diverge from the main topic of the

cluster [13].

Wayne Xin Zhao [14] et.al. suggested that while being Twitter as a new form of social media and containing useful information, content analysis on Twitter has not yet been well premeditated. They compared the content of Twitter with a traditional news medium i.e. New York Times by unsupervised topic modeling. They used a Twitter-LDA model for discovering topics from a representative sample of the entire Twitter. With the help of text mining technique they compared Twitter topics and the topics from New York Times while considering the topic categories and types. They presented a study based on the relation between the magnitude of prejudiced tweets and retweets and also over topic categories and types. With the help of Twitter-LDA model analyzing short tweets showed its efficiency compared with existing models. They proposed the concept of topic categories and topic types facilitating analysis of the contemporary differences between Twitter and they found that Twitter is a good source of entity oriented topics having low coverage in traditional news media. While Twitter users having low interests in world news, dynamically spread news of important world events [14].

3 CONCLUSION

The main aim of this paper is to analyze the various applications implemented for Online Social Media such as Filtering of message that are publicly posted on users wall, these wall messages may contains some unwanted or abuse words which needs to be filtered, Also the student's opinion their views on the wall and hence an opinion can be summarized from these views. Hence various technique and their limitations and advantages are discussed and analyzed here in the paper.

REFERENCES

- [1] J. S. Park, S. Kim, C. Kamhoua and K. Kwiat "Towards Trusted Data Management in OSN Services" World Congress on Internet Security, IEEE- 2012, 978-1 -908320-04/9 pp. 202-203
- [2] L. Kagal, T. Finin, M. Paolucci, N. Srinivasan, K. Sycara and G. Denker "Authorization and Privacy for Semantic Web Services" IEEE Intelligent Systems 2004, 1541 - 1672/04, pp. 50-56
- [3] J. Srivastava, P. Desikan and V. Kumar "Web Mining - Concepts, Applications and Research Direction", Chapter 21 pp. 399-417.
- [4] Soumen Chakrabarti, Byron Dom, David Gibson, Jon Kleinberg, S. Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, and Andrew Tomkins, "Hypersearching the web", Scientific American, 1999
- [5] Liaoruo Wang, Tiancheng Lou, Jie Tang and John E. Hopcroft "Detecting Community Kernels in Large Social Networks", 2011
- [6] Alan Mislove, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel and Bobby Bhattacharjee, —Measurement and Analysis of Online Social Networks || , ACM, 2007
- [7] Walter Willinger, Reza Rejaie, Mojtaba Torkjazi, Masoud Valafar and Mauro Maggioni, —Research on Online Social Networks: Time to Face the Real Challenges || , 2009
- [8] Xin Chen, Mihaela Vorvoreanu and Krishna Madhavan, "Mining Social Media Data for Understanding Students' Learning Experiences", IEEE Transactions, 2014.
- [9] Mattias Rost, Louise Barkhuus, Henriette Cramer and Barry Brown "Representation and Communication: Challenges in Interpreting Large Social Media Datasets", ACM, 2013.
- [10] Phil Long and George Siemens "Penetrating the Fog: Analytics in Learning and Education", Educause Review, 2011.
- [11] Roja Bandari, Sitaram Asur and Bernardo Huberman, "The Pulse of News in So cial Media: Forecasting Popularity", 2011.
- [12] Liangjie Hong, Ovidiu Dan and Brian D. Davison "Predicting Popular Messages in Twitter", ACM, 2011.
- [13] Hila Becker, Mor Naaman and Luis Gravano "Selecting Quality Twitter Content for Events", Association for the Advancement of Artificial Intelligence, 2011.
- [14] Wayne Xin Zhao, Jing Jiang, Jianshu Weng, Jing He, Ee Peng Lim, Hongfei Yan and Xiaoming Li, "Comparing Twitter and Traditional Media using Topic Models", 2011.