A Survey on Relation Between Social Media and Business Firms

[1]M. Srinivasa Rao, [2]B. Jayaram
[1]Associate Professor, [2]Assistant Professor
[1,2] Department of CSE, MLR Institute of Technology, Hyderabad – 500 043

Abstract:- Social media data mining is a playing very vital role in individual life and business firms for their business purposes also. In this paper I have presented a few social networking websites with its growing amount of users rapidly. Architecture for processing big data is also discussed. Relation between business firms and social media is also discussed with a sample scenario. Apart from those challenges in dealing with social media is also discussed.

Keywords: Social Media, Data Mining,

1) INTRODUCTION:

Social big data networking is a collection of various types of data from various web sources (E.g.,: Facebook, twitter, LinkedIn, Wechat, Whatsapp etc) used for social data mining. The basic 5V properties [5], [1] of big data volume, velocity, variety, value, and veracity, which make it difficult to handle such big data using traditional tools and techniques. So various tools came into existence using cloud computing Social big data networking is a collection of various types of data from various web sources (E.g.,: Facebook, twitter, LinkedIn, Wechat, Whatsapp etc) used for social data mining. The basic 5V properties [5], [1] of big data volume, velocity, variety, value, and veracity, which make it difficult to handle such big data using traditional tools and techniques. So various tools came into existence using cloud computing Social big data networking is a collection of various types of data from various web sources (E.g.,: Facebook, twitter, LinkedIn, Wechat, Whatsapp etc) used for social data mining. The basic 5V properties [5], [1] of big data volume, velocity, variety, value, and veracity, which make it difficult to handle such big data using traditional tools and techniques. So various tools came into existence using cloud computing.

The top tools [1] in the market which are currently used for processing big data are Hadoop[6], HD Insight, NoSQL, Hive, Scoop, Poly Base, MsExcel, Presto, MangoDB. These tools are used to process the big data based on the real time application and size of data.

A majority of those tools find difficulty in classifying the data and its types and the real time application where it is used for processing. The social media data processing can be done based on user opinions and interest. Since every user opinion varies with other user so processing social big data is a challenge.

General profile-based communication and networking of people via the Internet is done in all social data mining websites. Here the main focus is on disclosing user profile features and collaborative exchange of information within a larger community of users [7].

Some of the leading social networking sites [10] as on January 2018 with its list of users in millions are shown below in figure 1 and it is compared with chat applications in existence.
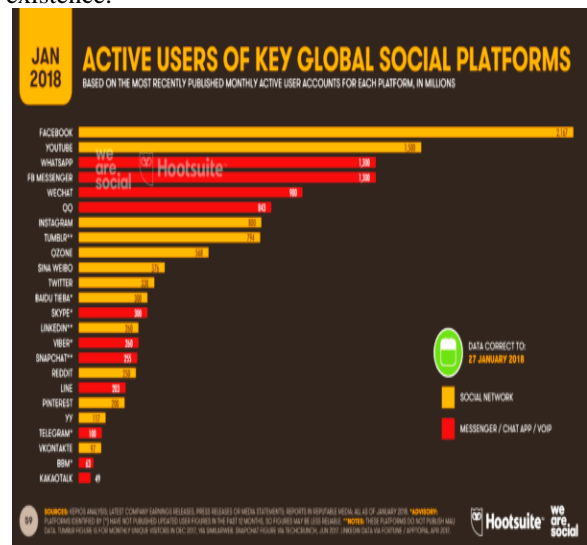

Figure 1: The number of users in social networking websites.

From figure 1[11], we can come to a conclusion that people prefer social media platforms more than chat applications since it can be used for multiple purposes compared to traditional applications.

The major point for discussion here is: What do social media users think about social media data mining [8]?

- Diversity in initial responses to scenarios.
- Diversity in group discussions,
- Obtaining diversity through fairness.

The properties of Social Influence[9] based on the above mentioned factors are listed below:

- Dynamic.
- Propagative.
- Composable.
- Measurable.
- Subjective.
- Asymmetric.
- Event sensitive.

From these we can plot the relationship between social influence analysis and big data [9] as represented in fig 2, which includes the following parts:

- Cloud storage for storage of data using cloud computing techniques.
- Big Data Preprocessing.
- Social influence analysis.
- Application of social influence analysis.

Here the 3rd part, social influence analysis can be classified into broad categories [9] as shown in fig 3 with its classification.
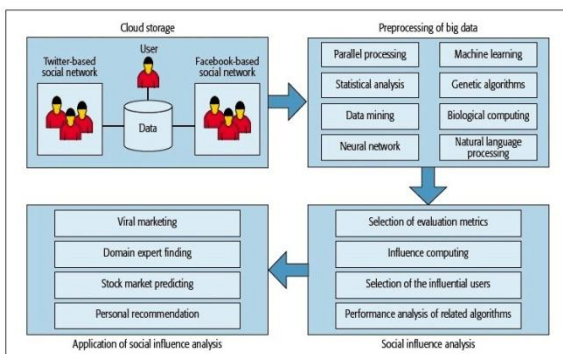

Figure 2: Relationship between social influence analysis and big data

Data collection: [9] from social networks: It is very critical phase in any data mining perspective. It requires the availability of data from social networking websites where data is mainly generated by online shopping, advertising, instant messaging, and mobile communication, the collection of raw data from online sources (e.g., Twitter, Facebook)

Data preprocessing: [9] In order to improve the performance and convenience of processing, we need to perform data preprocessing where we remove irrelevant information from social influence analysis. Privacy protection should also be taken care in this phase.

Selection of evaluation metrics [9]: To extract a set of evaluation metrics it is very important to accurately characterize the characteristics of each user. And these evaluation metrics would be helpful to quantify the social influence of each user and to easily find the most influential top-k nodes.

Modeling and computing social influence [9]: Evaluation model and computing equations can be provided from the extracted evaluation metrics to a specific social network. Thus, the social influence of each user can be computed by collecting real-world data sets and integrating the computing equations.

Selection of the most influential top-k nodes [9]: The influence maximization algorithm is designed to find the most influential top-k nodes.

Performance analysis [9]: Simulation is done to validate performance (e.g., influential range, computational complexity) of the proposed algorithms based on a specific propagation model. For a good performance of algorithm low complexity is always preferred.

1.1 Why should business firms concentrate on social media advertising?

Initially many business concerns were using traditional types of communication and when social media was introduced majority of the business firms were trying to answer the following questionnaire.
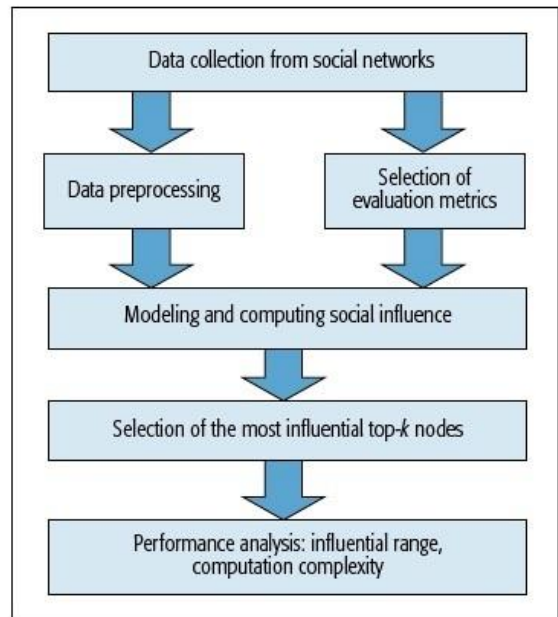

Figure 3: Architecture of social influence analysis

- What is social media?
- How to use social media?
- Why social media is to be used for business?
- What is the impact of social media on customers?

When comparing traditional media and social media [7], traditional media provides a one-way experience only (Eg: Advertising in TV media). Social media provides a two-way experience of interaction where consumers can interact with business concerns directly. Thus business through social media became very effective in current day scenario.

Social media data mining can be categorized according to user preference as discussed earlier. For example assuming the purpose of business as advertising, the top social networking websites for advertising [12] are shown below in figure 4. This data is taken as on April 2018.
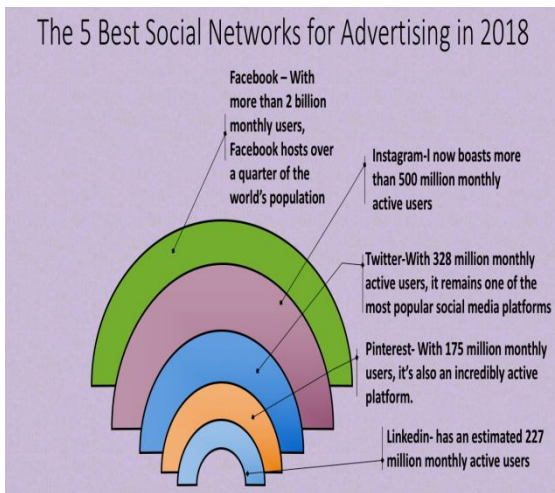
Figure 4: Top 5 social networking websites used for advertising.

2. CHALLENGES IN PROCESSING SOCIAL INFLUENCE ANALYSIS

There are various types of data and large volumes of data are to be processed from social media. So deciding a efficient algorithm always plays a major part for any processing. The following challenges are met in processing social influence analysis [9] in social media websites.

2.1 Determining Effective Evaluation Metrics

Due to heterogeneous of big data available and various algorithms designed choosing an effective metric for processing the big data through a effective tools always remains a challenge. Evaluation metrics is developed into a very important research area in data mining to address the challenges according to the real time application.

2.2 Considering the characteristics of dynamic evolution in large scale social networks.

The major important characteristic when dealing with social media data is that it is dynamic. So processing the data the algorithm or the tool should be selected accordingly.

The other important characteristics are growth of user in the social network is also dynamic. So it is also important to maintain a proper tree like structure or any other proper structure to process data.

2.3 Characterizing relationship in large scale social networks

Relationship has to be characterized for processing social data mining. Since every user can connect to multiple user through a relation. The relation might be like data fetched from blogs, newsfeeds, events etc. Researchers are recently using transfer entropy [13] to solve the problems related to relationship.

2.4 Guarantee of efficiency and scalability

Guarantee of efficiency in terms of social data mining is very difficult to predict the efficiency of the tools which are in use such open source tools like Hadoop [6], MangoDB etc.

Scalability is a process of increasing the networks by addressing at a macro level using various theory of computation algorithms such as NP-Hard and NP-Complete algorithms, Greedy algorithms. In general scalability and efficiency are indirectly proportional to each other. Hence the efficiency hence reduces with increase in scalability.

2.5 Determine the type of influence

The type of influence on the user can be broadly classified into three categories

- Positive influence.
- Negative influence.
- Controversy influence.

Positive influence gives trust for the user and also among relation of their friends and so on. So this increases the trust for group in total.

Negative influence indicates that there is no trust with user. And hence it leads to disagreement between connections of users. This also leads to rejections in relationship.

Controversy influence is normally a link between positive and negative influence. It indicates that opinion of one user varies from other user. So at times due to this varied opinion it may lead to controversy influence.

2.6 Considering the large collection of data for processing.

Initially in olden days traditional processing of data was in existence. Hence it created lot of problems in maintaining data and processing it. So it needed high time for processing.

These drawbacks are overcome by using automatic processing using various big data tools such as Hadoop [6], MangoDB, MsExcel, NoSql etc. Each has a specific structure for processing data. For social data mining mainly these tools are used to identify the purpose of the social media websites and also maintain privacy in data, which is also the most important topic in research these days.

2.7 Evaluating the influence of heterogeneous social networks

This is one of the important challenges in social networks. This includes the following challenges.

- Increase in social complexity involving nontrivial computation cost.
- Each type of entity is usually associated with one social network but participates in many networks with domain specific semantics.
- The information flow between two social networking websites may be bidirectional. So differentiating them is difficult.
- Since multiple social networks may be involved in multiple domains it is challenging to integrate them with a unified mechanism.

CONCLUSION

In this survey paper the reader gets insight regarding the importance of social media and its need for efficient data processing from social media. The efficient processing of big data requires various set of algorithms and cloud computing tools.

FUTURE IMPROVEMENTS

The future scope of this paper is choosing a particular real time application and process the data set with any cloud computing tool or algorithm. There are already so many algorithms for processing, choosing a good algorithm and getting an efficient processing for the business scenario should also be considered.

REFERENCES

[1] http://bigdata-madesimple.com/top-big-data-tools-used-to-store-and-analyse-data/
[2] Gema Bello-Orgaz, JasonJ.Jung, David Camacho," Social big data: Recent achievements and new challenges", Elsevier Publication August 2015.
[3] MohammadNoor Injadat, Fadi Salo and Ali Bou Nassif, Data Mining Techniques in Social Media: A Survey, Neurocomputing, http://dx.doi.org/10.1016/j.neucom.2016.06.045, June 2016
[4] Adil Fahad1, Najlaa Alshatri, Zahir Tari, Abdullah Alamri, Ibrahim Khalil, Albert Y. Zomaya2, Sebti Foufou, And Abdelaziz Bouras,"A Survey of Clustering Algorithms for Big Data: Taxonomy and Empirical Analysis" IEEE Transactions on EMERGING TOPICS IN COMPUTING, June 2014
[5] https://en.wikipedia.org/wiki/Big_data
[6] https://en.wikipedia.org/wiki/Apache_Hadoop
[7] MMF Naja, MII Mohamed, "Analysis of Systematic Data Mining Approaches for Achieving Competitive Advantage by Monitoring Social Media" in Kotelawala Defence University Sri Lanka Electronic Repository (KDU-ER) Pages 1- 7.
[8] Helen Kennedy, Dag Elgesem, Cristina Miguel, "On fairness: User perspectives on social media data mining" in Convergence: The International Journal of Research into New Media Technologies, Downloaded from con.sagepub.com at University of Sheffield on December 11, 2015
[9] Sancheng Peng, Guojun Wang, and Dongqing Xie, "Social Influence Analysis in Social Networking Big Data: Opportunities and Challenges" IEEE Network January/February 2017
[10] https://www.google.co.in/search?biw=1366&bih=662&tbm=isch&sa=1&ei=OP0pW-m7NYfevATBuquoBg&q=Leading+Social+Networking+Platform+2018&oq=Leading+Social+Networking+Platform+2018&gs_l=img.3...25899.30163.0.30509.13.12.0.0.0.0.485.1643.2-3j1j1.5.0....0...1c.1.64.img..8.0.0....0.Dq5FhbKLwms#imgrc=gJS-nz8DeoR6jM:
[11] https://www.google.co.in/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwj0oZOG6OHbAhVKpI8KHbUBCdUQjRx6BAgBEAU&url=http%3A%2F%2Fmarketingtrendz.in%2Ftop-5-social-media-platforms-you-should-invest-in-2018%2F&psig=AOvVaw1kwpYdMHf47emFeUHd_pvt&ust=1529564890764451
[12] https://www.google.co.in/search?hl=en&tbm=isch&source=hp&biw=1366&bih=662&ei=5DQrW8zrEoOavQT-jYHIDQ&q=the+5+best+social+networks+for+advertising&oq=the+5+best+social+networks+for+advertising&gs_l=img.3...1016.24907.0.25185.50.20.3.27.0.0.1372.4849.0j5j6j2j1j7-1.15.0....0...1ac.1.64.img..5.13.4232...0j0i8i30k1j0i24k1.0.uPpf6fgtInI#imgrc=Y9urKHtvt4cFGM:
[13] S. He *et al.*, "Identifying Peer Influence in Online Social Networks Using Transfer Entropy," *Proc. Pacific Asia Workshop Intelligence and Security Informatics*, Beijing, China, Aug. 2013, pp. 47–61.
[14] Mariam Adedoyin-Olowe, Mohamed Medhat Gaber and Frederic Stahl , "A Survey of Data Mining Techniques for Social Media Analysis"
[15] Mariam Adedoyin-Olowe1, Mohamed Medhat Gaber and Frederic Stah, "A Survey of Data Mining Techniques for Social Network Analysis"

BIBILIOGRAPHY

Mr. M.SrinivaaRao, is currently working as Associate professor in MLR Institute of Technology, Hyderabad. He has around 20 years of experience in teaching and pursuing PhD in network security. His areas of interest include data mining, cryptography, machine learning and image processing.

Mr. B.JAYARAM is currently working as Associate professor in MLR Institute of Technology, Hyderabad. He has around 12 years of experience in teaching. His areas of interest include data mining, cryptography, machine learning and image processing.