UGC Approved Journal

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified

Vol. 4, Issue 9, September 2017

Case Study of Social Network Data Mining and Analysis

Suryakumar B[1], Dr. E. Ramadevi[2]

Ph. D Research Scholar, Department of Computer Science, NGM College, Pollachi, India[1]

Associate Professor, Department of Computer Science, NGM College, Pollachi, India[2]

Abstract: Social networking has been archived remarkable importance in the world. Social network sites like Twitter, Facebook, Lnkedin and google plus gained remarkable attention in last decades. The people are depending social medias for information gathering, news, and openion polls of other users in different subjects. They gathering massive data from social network sites, it causes them to generate big data chararistised by different types of computational uses namely noise, size and dynamism. These of the issues make social data gathering process is very complex to analyse. It is a tedious task analysing them manually. It results, data mining provides wide ranges of systematic techniques for sensing useful knowledge from huge data sets of trends, patterns and rules[1].The data mining technology are used for information gathering, retrival machine learning statistical modelling. The major techniques employed in the processing and analysis of data interpretation process in the cources of data analysis process. In this case study discuss different kinds of data mining techniques used in social media mining diverse aspects of the social network over the decades going from the historical data gathering techniques to the up-to date models, in corperating over dominent technique named TRCM.

Keywords: Social Network, Social Network Analysis, Data Mining Introduction.

I. INTRODUCTION

The social network is used to describe web based services that help us to create a public, semi-public profile with in a domain such that can connect communicate with other users within the network. Social media network has improved in the concept of web-2 technology by enabling exchange and formation user generated content. Social network is graph consiting of nodes and link used to represent the relation of network sites. The nodes contains entities and relationship between them results the links (presented in Figure1).

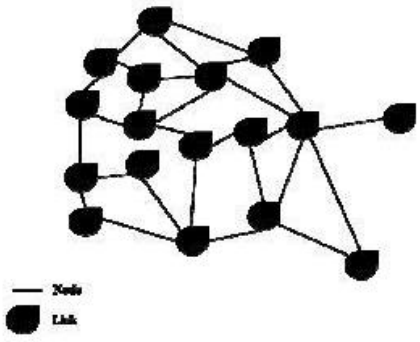

Fig. 1. Links and nodes in social Network sites

Now a days social networking are the important sources on the intractions and content sharing. It is based on subjectively observations, influences, assessments, expressions exicuted as in text, reviews, blog pages, news, reacting remakes or some other documents [2]. In the past decades before the advent of social network the home page was popularly used for share information in internet. After the invention of social network media enables a rapid information exchange between users regardless of the graphical locations. Most of the individual, organisations, even state governments now follow the social network. The social network sites empowers big organisations, government officials and celebrities and government bodies to gather knowledge have their audience responses to postings that depends them out of the enormous data generated from the social network sites(as shown in Figure 2). It helps the effective collection of large scale data which enables to major computational problem solving process. The efficiency of the data jmining techniques have been comparable of handling three prominent disputes namely noise, size,

UGC Approved Journal

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified

Vol. 4, Issue 9, September 2017

dinamism. The nature of social network data sets automated information processing for analysing it with in a limitted time period. The mining techniques also require huge data sets to mine remarkable pattern of data.

Figure 2 Data generated every minute in social media network

In this case study section 2 analyses social network background. Section 3 describes research issues on social network analysis. In the section 4 covers the mining tools used for graph thereoratic tools. Section five includes the overview of tools use to analising options. Section 6 presents some of the sentiment analysis technique. Section 7 deals with un-supervised classification of data.Section 8 includes topic detection process. The case study section 9 indicates the direction for the future work.

II. BACKGROUND OF SOCIAL MEDIA NETWORK

The reason of attraction of social media adiction of ordinary people is due to its free of cost availability and real time they can react to postings. Some situations users of social media make decision based on information hosted by unfamilier individuals on social network [3]. Increasing the degree of reliance on the credibility of social media sites. The speciality of social network site is it also given its users the privilage to give their oppenions with very little or without any restrictions.

III. POWER TO THE USER NEED

Most common factor of social media sites which have undoubtedly underlines unimaginable previlage or right on their users to access readly avilable and unsencered informations. For example Twiter permits user to post many kinds of events in real time way. It will very better method for broadcasting purposes coparatively on the other ther traditional news media. The speciality of social network media is allow the user to express their openion, views, be postive or negative[4]. The business people and organisations are very aware of the significance of consumers openion posted on social network media. The consious of the consumer feed back is very essential for their important of their business prosperity. The important personalities such as celibrites and goverment officials even politicians are being very concious how they are percived on social network. They arevery vigilant for knowing the how auidence react to the issues that concerns them[5][6][7].

The background of social network sites during the current business era social network sites have very popular affordable and universaly accepted information exchange media mean that has pay an important roll in making globe as village. The common use of social networking sites are for information exchange personal activities sharing, online media sharing, product reviews etc. The another important use of social media are professional profiling and advertisements it is also using for openion poles and sentiment exapression. Now a days social media are using live intraction to viewers posting news alerts, political debates and breakiing news. Due to enourmous amout of data being generated on social network it is a tedious task to find a computational means of categorising, classifing, filtering and analysing the social network data.

IV. MAJOR RESEARCH ISSUES

Followings are the common research issues found on social network analysis are identified are listed below.

UGC Approved Journal

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified

Vol. 4, Issue 9, September 2017

1. Structural and linkage based analysis

This is an analysis of structural level and linkage behaviour of social network sites. This analysis based on relevant nodes ,links, communities and imminent areas of social network(aggrwl-2011).

2. Statistic and dynamic Analysis

Dynamic analysis of stram based network like facebook and yutube are very difficult to carry out . The flow of data on these network are high speed and high capasity. The Dynamic analysis these networks are shown in the area of interaction between entities. In the case of statistical analysis bebiliographic network is permits to be easier to carry out than those in streaming networks statistic analysis persumed that social network changes gradulay over time analysis on the entire network can be done in batch mode.

3. Recommender System in Social Network Community

Based on the mutuality between nodes in social network groups,**collaborative filtering (CF)** technique, which forms one of the three classes of the **recommender system (RS)**, can be used to exploit the association among users . Items can be recommended to a user based on the rating of his mutual connection. Where CF's main downside is that of data sparsity, **content-based** (another RS method) explore the structures of the data to produce recommendations. However, the hybrid approaches usually suggest recommendations by combining CF and content-based recommendations. The experiment in] proposed a hybrid approach named **EntreeC**, a system that pools knowledge-based RS and CF to recommend restaurants. The work in improved on CF algorithm by using a greedy implementation of **hierarchical agglomerative clustering** to suggest forthcoming conferences or journals in which researchers (especially in computer science) can submit their work.

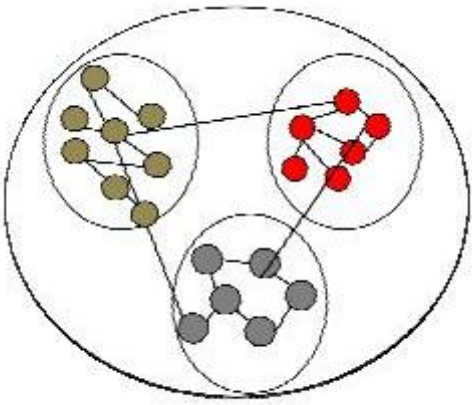

Fig. 4 Social Network Community Structure

V. SOCIAL NETWORK OPENION ANALYSIS

For the purpose of openion analysis verious methods have been developed to analyse the openion are derived from product,events, services or the personality reievew on social network[8]. As per the Technorect, about 75000 new blog sites and 1.2 Million news post giving on product and services are generated every day[9]. Every minutes a massive data is generated in social network sites it contains openion of users as regards diverse subjects ranging from personal to global issues[10]. Various data mining tools already used for openion and sentimental analysis include group of simple counting methods to machine learning. The process which categorising opening based test using binary distinction of positive aginst negatives[11].

1. Homophily clustering in openion formation

Openion of viewers on social network is based largly on there views and personal views and can not b e hold as absolute fact. Ocial media users that express that same openion are linked under the same nodes andthose oposing openion are linked in other node (As shown in Figure 5). This type of openion formation concept is refeered to as homophily in social network. This concept can also be demonstrate using other critieria such as race and gender[12].

2. Recomender system in social network media

The social nework group mutality between nodes are one important factor. Collaborative filtering technique which forms one of three class of the recomended system. RS can be used to demostrate the assosiation amoung users[13]. Here collaborative filtering techniques main limitation is that of data saparsity. In the condent based recomended

UGC Approved Journal

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified

Vol. 4, Issue 9, September 2017

system explores the structure of the data to procedure based recommendations. The hybrid approches sugests recommendation by combining CF and content based recommendations.

VI. SENTIMENTAL ANALYSIS OF SOCIAL NETWORK

The concept of sentimental analysis has been reffered as to as discovery and recognisation of positive or negative expression of openion by people on the diverse subject matters of interest. The users express the openion are often convincing and these indicator can be used form the basis of choices and decisions made by people, certian products and services or views of political candidate during election[14][15].

1. Semantic web on social networks

The main fucntion of semantic web platform is knowledge sharing and reuse possible over different community edges and application. Semantic web enhanses the knowledge sharing of the prominence of semantic web community. The research work in [92] employed FOAF(Friend of a Friend] to explore how local and community level group develop and evolve large scale social network on the semantic web.

VII. UNSUPERVISED CLASSIFICATION OF DATA

An unsupervised learning algorithm can be used to rate a reveiw as thumbs up or thumb down[85]. This can be by expresing out phrases that include objective or addverb(part of speech tagging)[75]. The sematic onrentation of every phrase can be aproximated using PM-1IR[120] and then classify the reveiw using average semantic orientation of the phrase.

Semi-supervised Classification

Semi-supervised learning is a goal-targeted activity but unlike unsupervised; it can be specifically evaluated. Authors of [31] worked on a mini training set of seed in positive and negative expressions selected for training a term classifier. Synonym and antonym comparatives were added to the seed sets in an online dictionary. The approach was meant to produce the extended sets P' and N' that makes up the training sets. Other learners were mployed and a binary classifier was built using every glosses in the dictionary for both term in P'∪ N' and translating them to a vector. Their approach discovers the origin of information which they reported was missing in earlier techniques used for the task. **Semi-supervised lexical classification** proposed by [77] integrated lexical knowledge into supervised learning and spread the approach to comprise unlabelled data. Cluster assumption was engaged by grouping together two documents with the same cluster basically supporting the positive - negative sentiment words as sentiment documents. It was noted that the sentiment polarity of document decides the polarity of word and vice versa.

VIII. DETECTCTION OF TRAFIC AND TRACKING

The process of Topic Detection and Tracking (TDT) on social network exexicutes different techniques for discovering the emergent of new topics (or events) and for tracking their subsequent evolvements over a time period. The topic detection and tracking is receiving high level of attention recently. Many researchers and authors are conducting experiments on TDT on social network sites, especially on Twitter [1]; [37]; [9]; [10]; [63]; [70]; [58]. In the case of [25] support vector machine (SVM) was found to be efficient in training Twitter hashtags metadata when predicting the political alignment of twitter users. Authors of [9] used an incremental online clustering algorithm to cluster a stream of Twitter messages in real time.

They trained a Naïve Bayes-Text classifier to distinguish between fastest-growing real-world events contents and nonevents contents on Twitter. The performance of the training set shows the precision of all classifier computed in 10-fold cross-validation. The analysis in [11] used a range of query-building approaches to automatically enhance user-contributed information for planned events with robustly generated Twitter contents. Their approach used browser plug-in script and a customizable web interface to identify relevant Twitter content for planned events.

IX. CONCLUSION

In this case study covers different data mining techniques that have been used for social network analysis. The techniques exicuted from unsupervised to semi supervised learning methods. Till now different levels of results have been atchived either with solitory of combined techniques. The output of the experiment on social media analysis is belived have shed more positive outcomes on the activities and stucture of social network. The different kinds of experimental out come have also supports the relevance of data mining techniques in retriving information and condence from huge data generated on social network sites. The future of the case study tend to findout novel state of

UGC Approved Journal

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified

Vol. 4, Issue 9, September 2017

the art of data mining technique used for network analysis. This case study compares similar data mining toolsand recomended most appropriate toolsfor the large data sets to be analysed.

Table.1 List of Data mining Techniques currently in Used in Social Network Analysis.

Approach	Tools	Experiments	Authors/dates
Graph Theoretic	Centrality measure	Inspects representation of power and influence that forms clusters and cohesiveness.	Burts (2005) Borgatti & Everett (2006)
	Parameterized centrality metric	Studies the network structure and to rank nodes connectivity.	Ghosh & Leman (2011)
	a-centrality	Measures the number of alleviated paths that exist among nodes.	Bonacich & Lloyd (2001)
Community Detection (hierarchical clustering)	Vertex clustering	Measures pairwise length between vertices.	Papadopoulos et al (2012)
Opinion Analysis	Aspect-Based/Feature-Based	To identify positive or negative opinion sentences in product reviews.	Hu & Liu (2004)
Opinion Formation	Homophily Clustering	Used to link same opinion under the same nodes.	Lynn Smith-Lovin & Cook, (2001) Jackson M, 2010
Opinion Definition and Opinion Summarization	Support Vector Machine (SVM/linear kernel	Used to learn the polarity of neutral examples in documents.	Ku et al (2006)
Sentiment Orientation (SO)	hierarchical classification technique	Used to improve the performance of mood classification.	Keshtkar, & Inkpen (2009)
Product Ratings and Reviews	Matrix factorisation method	Used to increase rating predictions and estimate accurate strengths of trust associations	Au Yeung and Iwata (2011)

REFERENCES

[1] Kagdi, H., Collard, M. L., Maletic, J. I.: A survey and taxonomyof approaches for mining software repositories in the context of software evolution. J. Softw. Maint. Evol.: Res. Pract, 19, 77-131,2007.
[2] Liu, B.: Sentiment analysis and opinion Mining. AAAI-2011, San Francisco, USA, 2011.
[3] Pang, B. and Lee, L.: Opinion mining and sentiment analysis;Foundations and trends in information Retrieval; Vol. 2, Nos. 1–2,1–135, 2008.
[4] Aggarwal, C.: An introduction to social network data analytics.Springer US, 2011

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified

Vol. 4, Issue 9, September 2017

[5] Castellanos, M., Dayal, M., Hsu, M., Ghosh, R., Dekhil, M.: U LCI: A Social Channel Analysis Platform for Live Customer Intelligence. In: Proceedings of the 2011 international Conferenceon Management of Data. 2011

[6] Chen, Y., Lee, K.: User-centred sentiment analysis on customer product review. World Applied Sciences Journal 12 (special issue on computer applications & knowledge management) 32 – 38,2011. ACM, New York, NY USA, 2011.

[7] Godbole, N., Srinivasaiah, M., Steven, S.: Large Scale Sentiment Analysis for News and Blogs. In: Proceedings of the International Conference on Weblogs and SM (ICWSM), 2007FLEXChip Signal Processor (MC68175/D), Motorola, 1996.

[8] Kim, P.: The Forrester Wave: Brand Monitoring, Q3 2006," Forrester Wave (white paper), 2006.

[9] Tepper, A.: How much data is created every minute?[INFOGRAPHIC]. 2012, http://mashable.com/2012/06/22/datacreated-every-minute/. Retrieved on 16/102013 at 19.00

[10] Goldberg, A., and Zhu, X.: Seeing stars when there aren't many stars: Graph-based semi supervised learning for sentiment categorization. In HLT-NAACL 2006 Workshop on Textgraphs: Graph-based Algorithms for Natural Language Processing, 2004.

[11] Jackson, M. O. Social and economic networks. Princeton University Press, 2010

[12] Liu, F., Lee, H. J.: Use of social network information to enhance collaborative filtering performance. Expert Systems with Applications, 37, 4772-4778, 2010.

[13] Kaschesky, M., Sobkowicz, P., Bouchard, G.: Opinion Mining in Social network: Modelling, Simulating, and Visualizing Political Opinion Formation in the Web**.** In: The Proceedings of 12th Annual International Conference on Digital Government Research, 2011.

[14] Pang, B. and Lee, L.: Using very simple statistics for review search: An exploration," In: Proceedings of the International Conference on Computational Linguistics (COLING) (Poster paper), 2008.

BIOGRAPHY

B. Suryakumar received M.Sc and MCA degrees in Computer Science from Annamalai University, Tamilnadu. Currently he is doing PhD in Computer Science at Bharathiar University, Coimbatore. His research interest lies in the area of Data Mining, Networking and Data security.

Dr. E. RamaDevi received PhD degree in Computer Science from Mother Teresa Women's University, Kodaikanal. Currently she is an Associate Professor in Computer science at NGM College, Pollachi, India. She has got 14 years of research experience and has more than 19 years of teaching experience. Her research interest includes areas like Data Mining, Knowledge base System, Intelligent and Control System and Fuzzy Logic. She has presented various papers in national and International Conferences and published 10 research papers on refereed journals.