

Analysis of Data Mining Methods for Social Media

Keshav S Rawat

Department of Computer Science & Informatics, Central university of Himachal Pradesh

Dharamshala (Himachal Pradesh)

Email:Keshav79699@gmail.com

Abstract: Due to increasing and rapidly use of social network on the World Wide Web, explosive growing of peoples willingness to exchange their opinion and interest on various social media like – face book, twitter, blogs etc. These web sources contain unstructured data in the field of business, government, education, medical and health. The objective of this paper is to analyze the various data mining techniques that are utilized by these web sources. After analysis of data mining method, we found that graph partition method and sentiment analysis are the good techniques to mining of web social sites data.

Keywords- *Data mining, social media, sentiment analysis, opinion analysis.*

1. Introduction:

Now a days, World become a small town due to growing use of web and social media , one can who belongs to different part of world may easily share his opinion, emotions, videos, pics, feelings etc by social media and now product is being promoted, advertized by these social sites too. Social media play a very important role in e-learning, medical science, tourism, business etc to get benefit in terms of knowledge and profit.

Social networking are electronic and portable based computer plate form that permit client produced criticisms, supposition, thoughts, interests, pictures/recordings to be made and exchanged. unstructured content, online journals, dialog gatherings, wikis and news, all commonly yielding data available through the web. To increase learning from social media flow a few strategies to separate data representative.

The first social media networking site was introduced in 1994 by geo cities that allowed user to create and share their home pages. The first social networking website was sixdegree.com, which was introduced in 1997[1]. The social site data is unstructured and unorganized that may be in form of image, text and other multimedia forms [4] and it also contains real time data which is not deal by traditional data mining techniques, hence data mining methods are very important issue for knowledge extraction from web social media[5].

The remaining paper is containing following sections: Section 2 describes the data mining process. Section 3 described data mining techniques for social media. Section 4 describes Comparison of data mining techniques. Section 5 contains the Conclusion and last section of the paper contains references.

2. Data mining and Knowledge generation

Data mining is the procedure of extraction of unstructured data from various web contents including social media[2]. The general point of the data mining process is to extract data from large datasets and change it into reasonable structure for additionally utilize [3]. The process of data mining is shown in fig 1. The process contains following steps – (i) Selection (ii) Preprocessing (iii) Transformation (iv) Extraction (v) Evaluation.

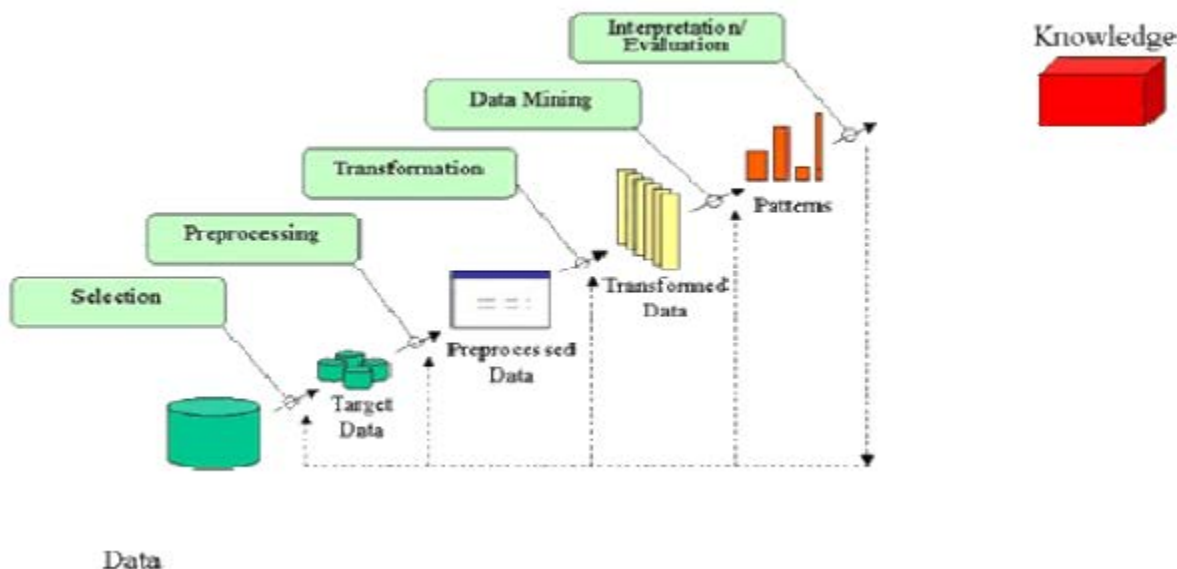


Fig 1: Data mining steps

The application areas of data mining are (i) Education (ii) Business (iii) Medicine (iv) Health (v) Finance (vi) Social Network etc.

3. Data mining techniques for social media

There are various data mining techniques are used for extracting knowledge from unstructured data-

- (1) Classification- It assigns data in a collection to object classes. Some popular classifications are (i) Decision trees(DT), (ii) ANN (iii) Rules Induction, (iv) Bayesian classifiers(BN), (v) SVM etc
- (2) Clustering- It is an unsupervised classification where data set is split into two groups call clusters so that same data come together in one cluster.
- (3) Association- It extract related correlations, patterns, and associations from sets of data items in the transaction databases.

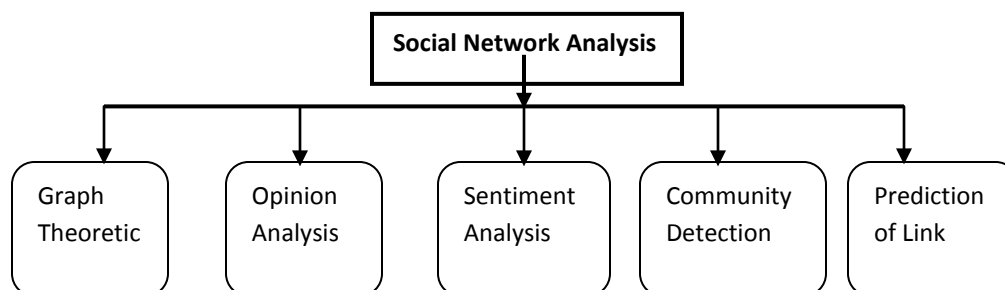
Some of the clustering techniques that are playing an important role in classification of social data [6] are described in table 1.

Table1: Taxonomy of clustering methods

Clustering Algorithms			
<i>Partitioning Based</i>	<i>Hierarchical Based</i>	<i>Density Based</i>	<i>Grid Based</i>
K – Means K – medoids K – modes PAM CLARANS CLARA FCM	BIRCH CURE ROCK Chameleon Echidna	DBSCAN OPTICS DBCLASD DENCLUE	Wave Cluster STING CLIQUE OptiGrid

Figure 2 shows various research issues for Information mining from social media sites –

(i) Graph Theoretic (ii) Opinion Analysis (iii) Sentiment Analysis (iv) Community Detection (v) Prediction of Link

**Fig 2: Research issues of Social media**

(1). Graph Mining- This method used graph theory to analyze large data set of logs of social media sites[1]. This method is very useful to find follower and influencers of network.

(2) Opinion Analysis- it refer to computation or evaluation of opinion , emotion and sentiment available on social media in form of text, pics, videos, blogs and discussion forum etc [7]. Opinion analysis is generally divided in two forms - (i) Feature Based Opinion Mining (ii) Corpus Based Opinion Mining

(3) Sentiment Analysis – This method is used to find or uncovers the user sentiment – positive or negative expression of opinion on matter available on social media sites. It classifies data available on social sites based on their polarity- positive or negative or neutral type. The mining system based on opinion is also referred as “Document Based Sentiment Oriented System” [7].



Fig 3: sentiment analysis steps

Sentiment analysis process is described in fig 3. It contains the following steps- (i) Opinion or review (ii) Subjective detection (iii) Feature selection (iv) Classification and extraction (v) sentiment analysis in terms of polarity.

(4) Community Detection- It is focused on the link part of networks over content of social site. Newman & Girvan's community search algorithm is most popular community detection algorithm [8].

(5) Link Mining- It is used to predict number of link , type of link between data items, reference, sub graph pattern etc[9].

4. Comparison of data mining techniques

In this paper, we analyzed various data mining techniques used in social networking sites and this can help to extract knowledge from social sites to give benefit to increase organization performance. The comparison of data mining techniques for social networking is shown in table 2.

Table 2: Comparison of mining techniques used in social networking

S. No.	Mining Approach	Method	Features
1.	Graph Mining	Graph partitioning method	Identify follower and influencers of network
2.	Opinion Analysis	Feature based and corpus based mining method	Evaluate emotions good or bad about review
3.	Sentiment Analysis	Machine learning approach	Classify opinion polarity – positive, negative or neutral
4.	Community Detection	Newman & Girvan's community search algorithm and Multi-scale detection	Determine densely connected node in network and scalable
5.	Link Mining	Link Prediction algorithm	Determine various patterns available in network

5. Conclusion & future work

In this paper we have compared various data mining methods for extracting information from the sources available at social media sites. This paper also discusses the some important clustering techniques for classification of data. Overall graph partition method and sentiment analysis are most suited for mining of data from social media sources. Future scope of this paper is to do detail comparative analysis of machine learning algorithm and tools for mining data from social media.

References-

- [1] Reza Zafarani, Mohammad Ali Abbasi, Huan Liu, "Social Media Mining: An Introduction", Cambridge University Press New York, NY, USA ©2014.
- [2] Hand David, Mannila Heikki, Smyth Padhraic.: "Principles of data mining", Prentice hall India, pp.1, 2004.
- [3] Wei Fan and Albert Bifet. Mining Big Data: Current status, and Forecast to the Future. SIGKDD Explorations 14(2):1-5, 2012.
- [4]A.L.Kavanaugh,E.a.Fox,S.D.Sheetz,S.Yang,L.T.Li,D.J.Shoemaker,etal.,Socialmediausebygovernment:fromthero utinetothe critical, Gov. Inf. Q29(2012)480–491
- [5]H.Chen,R.H.L.Chiang,V.C.Storey,Businessintelligenceandanalytics:frombigdatatobigimpact,MisQ36(2012)1165 –1188.
- [6] FAHAD, A; Alshatri, N.; Tari, Z.; Alamri, A; Y.Zomaya, A; Khalil, I;Foufou, S.; Bouras, A, "A Survey of Clustering Algorithms for Big Data: Taxonomy & Empirical Analysis," Emerging Topics in Computing, IEEE Transactions on , vol.PP, no.99, pp.1,1
- [7] Richa Sharma et al.," Opinion mining of movie reviews at document level", International Journal of Information Theory (IJIT), (IJIT), Vol.3, No.3, July 2014.
- [8] David F. nettleton, "Data mining of social networks represented as graphs", Expert system with Application, Oct 2012.
- [9] Lise Getoor, " Link Mining: A New data Mining Challenge", UMIACS, 415- 444,Volume 4, Issue 2,2013

