ISSN:0975-9646

Survey on intelligent Data Mining of Social Media for improving health Care

Aradhana. S.Ghorpade
P.G. Student, Department of Computer Science& Engineering, DYPCET, Kolhapur, Maharashtra, India

Abstract: In this paper we take into careful thought of the concepts used for algorithmic and data mining Perspective of Online Social Networks (OSNs). Few such factors include the availability of huge amount of OSN data, the representation of OSN data as graphs, and so on .the different data mining techniques and Limitation faced by this technique are discussed hence, this paper gives an idea about the key topics of using Data mining in OSNs which will help the researchers to solve those problems that still exist in mining OSNs. New approach is introduced for mining social media.

Keywords: Online Social Networks, Data Mining, neural networks, network based modeling.

I. INTRODUCTION:

Extracting knowledge from social media has currently attracted great interest in every field. Many popular OSNs such as Face book, Orkut, Twitter, and LinkedIn have become increasingly popular. Every community is using social media extraction. However, social media sites provide data which are vast, noisy, distributed and dynamic. Hence, data mining techniques provide researchers the tools needed to analyze such large, complex, and frequently changing social media data.

With the massive growth of social media (i.e., reviews, forum discussions, blogs and social networks) on the Web, organizations are increasingly using public opinions in these media for their decision making. Sentiment analysis or opinion mining is the computational study of people's opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes. The task is technically challenging and practically very useful. For example, businesses always want to find public or consumer opinions about their products and services. Potential customers also want to know the opinions of existing users before they use a service or purchase a product.

Our paper is organized as follows: Firstly explanation of various social media mining techniques. Then the proposed system and conclusion.

II. KEY RESEARCH ISSUES IN ONLINE SOCIAL NETWORK ANALYSIS

Some key research issues in mining social networking sites using data mining techniques [1] are discussed below:

1. *Community or Group Detection*

In general, community or group detection is based on study of the structure of the network and finding individuals that correlate more with each other than with other users. M. E. J. Newman, "Detecting community structure in networks [2] has reviewed algorithmic methods for finding common unities of densely connected vertices in network data.

2. *Link analysis*

In network theory, link analysis is a data-analysis technique used to evaluate connections between nodes. Connections may be identified among various types of nodes (objects), including organizations, people and transactions. L. Getoor and C. Diehl [3], performed survey on link mining. In which they have discussed that many datasets of interest today are best described as a linked collection of interrelated objects. These may represent homogeneous networks, in which there is a single-object type and link type, or richer, heterogeneous networks, in which there may be multiple object and link types (and possibly other semantic information). Examples of heterogeneous networks include those in medical domains describing patients, diseases, treatments and contacts, or in bibliographic domains describing publications, authors, and venues.

Link mining refers to data mining techniques that explicitly consider these links when building predictive or descriptive models of the linked data. This is an exciting, rapidly expanding area.

3. *Predicting Trust and Distrust among Individuals*

Due to the continuous expansion of communities in social networks, the question of trust and distrust among individuals in a community has become a matter of great concern. Past assessments reveal that some users try to either disturb or take undue advantage of the normal atmosphere of such online communities. A number of disciplines have looked at various issues related to trust.

One of the first works on this task was the *EigenTrust* algorithm [4] that aims to reduce the number of inauthentic file downloads in a P2P network. Guha et al. [5] proposed methods of propagation of trust and distrust, each of which is appropriate in certain circumstances.

4. *Behavior and Mood Analysis*

Discovering human behavior or human interaction based on data mining techniques is also an interesting research field that is gaining huge attention in research. Here, human behavior may indicate any human-generated actions such as clicking on a specific advertisement, accepting a friend's request, joining a group or discussion forum, commenting on an image, music, etc, or dating with a person, etc. Benevenuto et al. [6] measured the behavior of online social networks' users applying the proxy server-based measurement framework.

5. *Recommender Systems*

Recommender systems (RS) provider commendations to users about a set of articles or services they might be interested in. This facility in OSNs has become very popular due to the easy access of information on the Internet. Few important applications of RS are its use in several websites for recommendation of items such as movies, books, gadgets, etc. Recommender systems (RS) have developed in parallel with the web. A good survey on various RS can be found in [7].

6. *Influence Propagation*

Nowadays, as OSNs are attracting millions of people, the latter rely on making decisions based on the influence of such sites. For example, influence propagation can help decide which movie to watch, which product to purchase, and so on. Thus, influence propagation has become an important mechanism for effective viral marketing, where companies try to promote their products and services through the word-of-mouth propagations in OSNs. Domingos and Richardson [8] provided the first algorithmic treatment to deal with influence propagation problem.

7. *Expert Finding*

OSNs consist of several experts in a specific domain and other people who join the network to receive help from these experts. These OSNs can be used to search for such experts within a group of people. For example, a topic related expert can be searched based on the study of the link between authors and receivers of emails. Study on expert ranking algorithm is usually based on either domain knowledge driven methods or domain knowledge independent methods or both. The expert ranking problem is also researched on email communication relations [9].

8. *Opinion Mining*

OSNs have given rise to various review sites, blog repositories, online discussions, etc where people can express their ideas and opinions, exchange knowledge and beliefs, criticize products and ideas. Data mining of opinions on specific subjects allows the detection of user prospects and needs, and also feelings or reactions of people about certain beliefs, products, decisions or events. S. R. Das and M. Y. Chen, proposed an algorithm for sentiment extraction for small talk on the web" [10].

Extracting sentiment from text is a hard semantic problem. They develop a methodology for extracting small investor sentiment from stock message boards. The algorithm comprises different classifier algorithms coupled together by a voting scheme. Accuracy levels are similar to widely used Bayes classifiers, but false positives are lower and sentiment accuracy higher. Empirical applications evidence a relationship with stock values—tech-sector postings are related to stock index levels, and to volumes and volatility. A. Akay, A. Dragomir, and B. E. Erlandsson [11], developed a novel data mining method for extracting consumer opinion on diabetic disease based on the user post from the different forums.

Despite the extensive literature, none have identified influential users, and how forum relationships affect network.

III. PROPOSED WORK:

The proposed system will intelligently mine data from social media by using forum post and user feedback. Natural language processing and data mining techniques will be used for mining forum post and user feedback.

At first using the self-organizing maps (SOMs) exploratory analysis will be employed to assess correlations between user posts and positive or negative opinion on the medicine. Then it models the users and their posts using a network-based approach to find influential users.

Using the discussion of influential user the side effect of medicines will be identified which can be used to improve care.

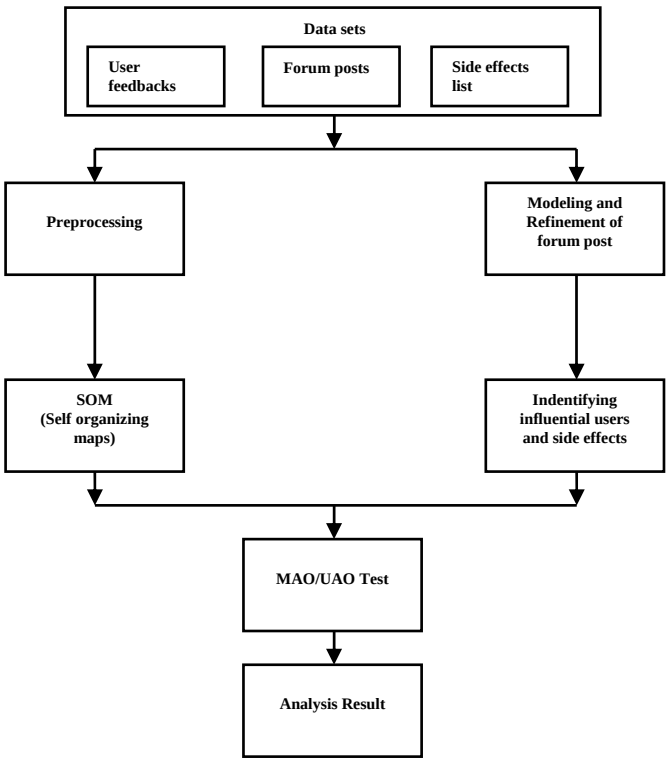

Fig: Proposed System Architecture

a) Data collection and preprocessing.

Data collection will be done from various sites, forums and feedbacks related specific medicine. Processing will be performed on the raw text data collected using the language processing libraries and algorithms to look for the most common positive and negative words and their term-frequency-inverse document frequency (TF-IDF) scores within each post. In previous study only post were used but to make the result more precise we are going to use the user feedbacks related to treatment and medicine.

b) Consumer sentiments analysis.

For this part of the analysis, all posts and feedbacks are labeled according to the general user opinion observed within the post and feedbacks as positive and negative before feeding the collected data for analysis via SOMs (self organizing maps). Subgroups (neurons) were formed on the basis of their weights assigned in the previous module. The similar weighted words are group in same neuron and network based model is formed. This neuron shows the positive and negative words correlation.

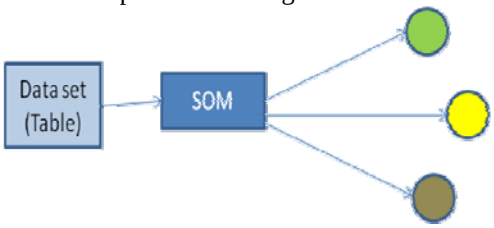

Fig: network based model using SOM.

c) Modeling forum posting:

Discovering influential users is the next step in our analysis. To this goal, we will build networks from forum posts and their replies. In a first step, we aimed at identifying *influential users* within our networks. Influential users are users which broker most of the information transfer within network modules and whose opinion in terms of positive or negative sentiment towards the treatment is 'spread' to the other users within their containing modules. To obtain this the previously derived algorithm will be used were [13] author proposed an approach in which transition probabilities for a random walk of length *t* (*t* being the Markov time) enable multi scale analysis.

d) Refining information module:

In the second step of our network-based analysis, we devised a strategy for identifying potential side effects occurring during the treatment and which user posts on the forum highlight. To this goal, we overlay the TF-IDF scores of the wordlist onto modules. The TFIDF scores within each module will thus directly reflect how frequent a certain side-effect is mentioned in module posts. The module average opinion (MOA) and user average opinion (UOA) is calculated.

e) Identification of side effects and performance analysis.

On the basis of the MAO and UAO the influential users are find out and only that modules are considered for further analysis. only the influenced users are analyzed and the common side effect related to medicine is obtained .further the T -Test can be applied to evaluate performance analysis of the obtained result .after the complete mining the obtained accurate result can be further use for many pharmaceutical companies and the user.

IV. CONCLUSION:

Social media can open the door for the health care sector in address cost reduction, product and service optimization, and patient care. The proposed work can be beneficial for in every sector to gain feedback of any product .new analysis can be added to make the system more refined. Healthcare providers could use patient opinion to improve their services. Physicians could collect feedback from other doctors and patients to improve their treatment recommendations and results. Patients could use other consumers' knowledge in making better-informed healthcare decisions.

REFERENCES:

[1] G Nandi1, A Das ," A Survey on Using Data Mining Techniques for Online Social Network Analysis", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 6, No 2, November 2013.

[2] M. E. J. Newman, "Detecting community structure in networks," *Eur. Phys. J.*, vol. 38, pp. 321–330, Mar. 2004.

[3] L. Getoor and C. Diehl, "Link mining: a survey," *SIGKDD Explore. Newsl.* vol. 7, pp. 3–12, Dec. 2005.

[4] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, "The eigentrust algorithm for reputation management in P2P networks", in the Proc. of the 12th International World Wide Web Conference, 2003, pp. 640–65.

[5] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, "Propagation of trust and distrust", in Proc. of the Int. Conf. on World Wide Web, 2004, pp. 403–412.

[6] F. Benevenuto, T. Rodrigues, M. Cha, V. Almeida, "Characterizing user behavior in online social networks", in Proc. of the 9th Int. Conf. on ACM SIGCOMM Internet Measurement Conference, 2009, pp. 49–62.

[7] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, "Recommender systems survey", In the Journal of Knowledge-Based Systems 46 (2013), pp. 109–132.

[8] E. Le Martelot and C. Hankin, "Multi-scale community detection using stability as optimization criterion in a greedy algorithm," *Proceedings of the 2011 International. Conf. erence on Knowledge Discovery and Information Retrieval (KDIR 2011), Paris, France: SciTePress, Oct. 2011*, pp. 216–225.

[9] S. Campbell, P. P. Maglio, A. Cozzi, and B. Dom, "Expertise Identification Using Email Communications", In Proc. of the 12th Int. Conf. on Information and Knowledge Management, 2003, pp.528-531.

[10] S. R. Das and M. Y. Chen, "Yahoo! for Amazon: Sentiment extraction from small talk on the Web," *Manag. Sci., vol. 53, pp. 1375–1388, Sep.* 2007

[11] Altug Akay (M'11), Network-Based Modeling and Intelligent Data Mining of Social Media for Improving Care.

[12] A. Akay, A. Dragomir, and B. E. Erlandsson, "A novel data-mining approach leveraging social media to monitor consumer opinion of sitagliptin," *J. Biomed Health Inform.* Vol: PP, Issue: 99.

[13] E. Le Martelot and C. Hankin, "Multi-scale community detection using stability as optimization criterion in a greedy algorithm," *Proceedings of the 2011 International. Conf. erence on Knowledge Discovery and Information Retrieval (KDIR 2011), Paris, France: SciTePress, Oct. 2011*, pp. 216–225.