NETWORK BASED MODELING AND INTELLIGENT DATA MINING OF SOCIAL MEDIA FOR IMPROVING CARE

[1]K.Sowmiya M.Phil., [2]Mrs.V.Parameswari   M.C.A, M,Phil.,

[1]Research Scholar, Department of Computer Science, Bharathiyar Arts and Science College for Women's, Deviyakurichi, Salem.

[2]Assistant Professor, Department of Computer Science, Bharathiyar Arts and Science College for Women's, Deviyakurichi, Salem.

Abstract

Social media provide an intelligence information extraction for the area of Health Informatics, including Bioinformatics, Image Informatics, Clinical Informatics, Public Health Informatics, etc. To improve the health care outcomes with reduced cost we propose the framework which concentrates on positive and negative symptoms of the disease and also the side effects of the treatment. Social media, assorting from individual messaging to live for as, is providing immeasurable opportunities for patients to converse their experiences with drugs and devices. Social media allows message, contribution, gathering information, and distribution in the healthcare space. Healthcare space is one which contains the information of the patients with their permissions. It provides an effective social networking environment. Social media provide an intelligence information extraction for the area of Health Informatics, including Bioinformatics, Image Informatics, Clinical Informatics, Public Health Informatics, etc. The proper way to mine information and drift from the knowledge is cloud. To improve the health care outcomes with reduced cost the system propose the framework which concentrates on positive and negative symptoms of the disease and also the side effects of the treatment. Many previous studies none have identified influential users, and how forum relationships affect network dynamics .In this project the system propose a two-step analysis framework. In the first stage the system employ exploratory analysis using the self-organizing maps. In the second stage the system model the users and their posts using a network-based approach.

1.   INTRODUCTION

Social network is a term used to describe web-based services that allow individuals to create a public/semi-public profile within a domain such that they can communicatively connect with other users within the network. Social network has improved on the concept and technology of Web 2.0, by enabling the formation and exchange of User-Generated Content. Simply put, social network is a graph consisting of node sand links used to represent social relations on social network sites. The nodes include entities and the relationships between them forms the links. Social media provide an intelligence information extraction for the area of Health Informatics, including Bioinformatics, Image Informatics, Clinical Informatics, Public Health Informatics, etc. To improve the health care outcomes with reduced cost we propose the framework which concentrates on positive and negative symptoms of the disease and also the side effects of the treatment. For this health care we take most prevalent cancers: lung cancer. Our proposed approach we employ investigatory analysis using the SOMs to check the correlations between user posts and positive or negative judgment on drug. Then we use the hierarchical clustering, it measures the starting with an empty network of n vertices and no edges, one add edges between pairs of vertices in order of decreasing similarity, starting with the pair with strongest similarity. Social media is one to extracting the information from internet.

Now a day it is used for extracting the data of patient's to know the experiences of patient symptoms. Social media, assorting from individual messaging to live for as, is providing immeasurable opportunities for patients to converse their experiences with drugs and devices. Social media allows message, contribution, gathering information, and distribution in the healthcare space. Healthcare space is one which contains the information of the patients with their permissions. It provides an effective social networking environment. The proper way to mine information and drift from the knowledge is cloud. Using network based analysis methods it model the social media such as Face book , Twitter, WebMD, etc.These analyses are made of nodes which may be individual or organization that connect the network in various associations such as interests, friendship, kinship, etc. Hence the representation of the information would be a graphical, which is convenient for the user visualization.

But in traditional level it uses the surveys to collect the data from the patients and which were resulting in small sample sizes per surveys and studies. With the help of social media, they were additional content is willingly available, particularly when shared with web-crawling and scraping software it allow real-time monitoring of modifications inside the network. In social network it provides an in-depth knowledge of social network dynamics. A network model could be used for replication studies of different network aspects such as mining how the users broadcast knowledge among themselves (news about epidemic or drugs' difficult effects). Other example is analyzing the augmentation of certain edges of networks and how certain information involves the enhancements (e.g., how certain user communities develop based on ordinary benefit about explicit diseases).In social media the duty of social networks involves data collection in difficult manner.

2. RELATED WORK

For this it used several techniques to employee, such as links, existence, estimation, object, group, and subgroup detection, and mining the data, link mining, classification through links, predictions based on objects. Other techniques are also based on the user feedback they were Link prediction, viral marketing, online discussion groups (and rankings) allow for the development of solutions based on user feedback. In social network, the data are heterogeneous, multi-relational, and semi-supervised, to gather these data is difficult. One method of extraction is link mining or relationship extraction, which provide the effective way for combining the social network using hypertext and web mining to extract data from web. Combining this, it involves classification Prediction; estimation, reconciliation, detection and finally extraction were done to extract the patient information.

There are two simple ways to mine this information: 1) crawling and 2) scraping, with crawling it afford site Application Programming Interfaces (APIs) then second rendered information from HTML pages. In our approach we employ investigatory analysis using the self-organizing maps (SOMs) to check the correlations between user posts and positive or negative judgment on the drug. Then we use the hierarchical clustering ,it measures the starting with an empty network of n vertices and no edges, one add edges between pairs of vertices in order of decreasing resemblance, initial with the pair with strongest resemblance.

We consequently improve the rescue modules with word regularity information from module-contained user's posts to obtain local and global measures of user's opinion and raise flag on potential side effects of Erlotinib, a drug used in the treatment of one of the most prevalent cancers: lung cancer. It provides a better result when compared to the previous work. Social media comprises interactive applications and platforms for creating, sharing and exchange of   user - generated

contents. The past ten years have brought huge   growth in social media, especially online social networking service s, and it is   changing our ways to organize and communicate. It   aggregates opinions and feelings of diverse   groups of people   at low cost.   Mining the attributes and contents of   social media gives us an   opportunity to   discover social structure characteristics, analyze action pattern s qualitatively   and quantitatively,   and sometimes the ability to   predict future human related events. In this paper, we firstly discuss   the realms which can   be predicted with current social media, then overview   available predictors   and techniques   of   predict ion, and finally   discuss   challenges and possible future   directions. Social media are platforms that allow common persons to create and publish contents.   Two worldwide popular social media websites Twitter and Face book, demonstrate   its   explosive growth and profound influence. Both Twitter and Face book are in the top 10 most - visited websites in the world according to Alex   ranking. Face book has more   than 800 million   active users, and by March 2011, on Twitter, there were   about 140 million information pieces created and transferred   daily. There is other specialized social media that are focused on entertainment, sports, finance and politics. Since there are many users sharing their opinions and experiences via social media, there is aggregation of   personal wisdom and different viewpoints.

3. SYSTEM ANALYSIS
EXISTING SYSTEM

Text and structural data mining of web and social media (WSM) provides a novel disease surveillance resource and can identify online communities for targeted public health communications (PHC) to assure wide dissemination of pertinent information. The system present results from text classification experiments that compare relevancy signatures, which use local linguistic context, with corresponding indexing terms that do not. In two different domains, relevancy Signatures produced better Results than the simple indexing terms. The system describes a text classification algorithm that uses linguistic expressions called "relevancy signatures" to classify texts.The pervasiveness and ubiquity of internet resources provide individuals with access to many information sources that facilitate self-diagnosis and provide means for nontraditional bio surveillance; for example, one can combine specific disease symptoms to form search queries. The results of such search queries often lead to sites that may help diagnose the illness and offer medical advice. Recently, Google™ has addressed this issue by capturing the query keywords and identifying specific searches involving search terms that indicate ILI. Current data mining research focuses primarily on algorithms to discover sets of attributes that can discriminate data entities into classes, such as shopping or banking trends for a particular demographic group. These approaches are difficult when key concepts involve relationships between the data points.

DISADVANTAGES

▶ None have identified influential users, and how forum relationships affect network dynamics.
▶ Sentences that do not activate any concept nodes are effectively ignored.
▶ The existing system did not do very well.
▶ Traditional information extraction approaches do not work well over the highly informal, noisy, and ungrammatical text common in social media.
▶ It does not handle the extraction and aggregation of the rare content that important applications need to extract from high-volume streaming sources.

PROPOSED SYSTEM

In this project the system proposed a two-step analysis frame work. The system employ investigatory analysis using the self-organizing maps (SOMs) to check the correlations between user posts and positive or negative judgment on the drug. Then the system forms the users and their posts using a network-based approach. This enables us to find the possible partition using complete linkage. The SOM results the reviews with the above two steps along with inter social dynamic maps. To explore the survey dataset we have used Self organizing map (SOM). SOM also known is an unsupervised technique that is most often described in the language of artificial neural networks. SOM provides a way of representing multidimensional data in typically two or three dimensions. This process of reducing the dimensionality of sectors is based on a data compression technique known as vector quantization. In addition, SOM creates a network that stores information in such a way that any topological relation-ships within the training set are maintained. Hence, SOMs are useful for visualizing large data sets of high dimensionality.SOM is an unsupervised, competitive learning approach in which only one neuron 'wins' each training phase. There are no connections between the neurons in the input and output layers. However, they communicate with each-other via a neighborhood function. Our approach of summarizing and visualizing individual patient profiles turns will prove to be useful in this, as concurred by the domain experts. Computational perspective, we would like to examine the feasibility of using Self-Organizing Map as a means of extracting useful information from survey questionnaires.

This study also focuses on improving data interpretation by revealing possible associations between the tendency of item non response and participant background variables. In the present case, item non response does not refer to participants who fail to return the survey but to the ones who choose not to respond to all questions and are nevertheless included in the analyses.Large surveys have pointed that, although respondents and non-respondents in patient satisfaction surveys may differ according to several demographic and clinical characteristics, differences in satisfaction between them tend to be relatively small and non-respondents do not constitute a homogenous group. Many highly sophisticated statistical methods to handle the problem of missing responses are now available, and the use of them is becoming a standard. The following two sections will summarize expected predictors of patient satisfaction and then explain the SOM. The patient satisfaction section is necessary for understanding the slightly ambiguous nature of the concept. Patient satisfaction is affected not only by factors related to care receivers and care providers but also by factors related to ways it is defined and measured. The present research is a purely quantitative approach, and thus, qualitative studies carried out on patient satisfaction were excluded from our review. The SOM section will introduce general arguments for why the SOM was used in this study; some technical details necessary for understanding how the method works are also presented. Most data mining and knowledge discovery techniques, including the SOM, are data driven, rather than theory driven.

Advantages:

▶ Proposed model includes some advance detection technique of inter-social dynamics, which incorporates the ratings, relationships and rankings.
▶ This also provides the word snippet option for formal language input in the drug review forums.

► This helps to predict the list of terms related to the medical domain which are drug names and disease with their symptoms.

► This method is used to convert the forum posts into the directional network.

► Mapping consumer opinion on drugs, treatments and other medical services is the major aim of our current proposal.

4. SYSTEM MODULES

1. Preprocessing Dataset
2. TF-IDF
3. Filter Stop words
4. Side Effect

PREPROCESSING DATASET

► A Rapid miner data collection and processing tree was developed to look for the most common positive and negative words, and their term-frequency-inverse document frequency (TF-IDF) scores within each post.

► We initially up-loaded the data into the first component.

► The uploaded data was then processed in the second component using several subcomponents that filtered excess noise to ensure a uniform set of variables that can be measured.
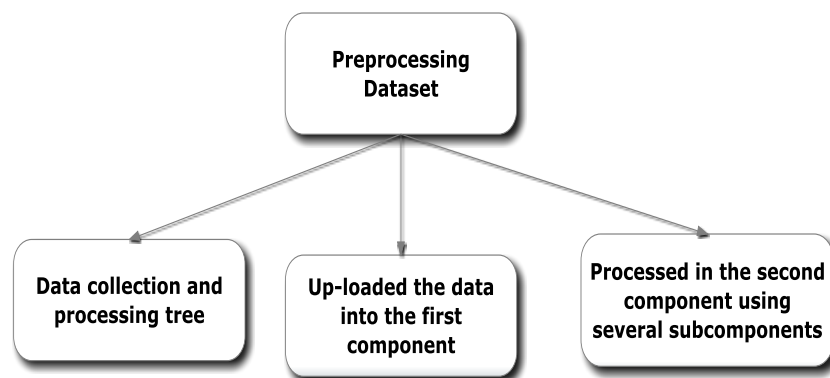

Figure 1: Flow of Preprocessing Dataset

TF-IDF

► Text data containing the highest TF-IDF scores were tagged with a modified NLTK toolkit to ensure that they reflected the negativity of a negative word and the positivity of a positive word in context.

► This approach was used before using negative tags on positive words.

► Rapid miner processing tree: we kept the side effect words with the highest TF-IDF scores.

► After these preprocessing steps, our forum data was represented as two sets of vectors containing the TF-IDF scores of the words in the two wordlist.
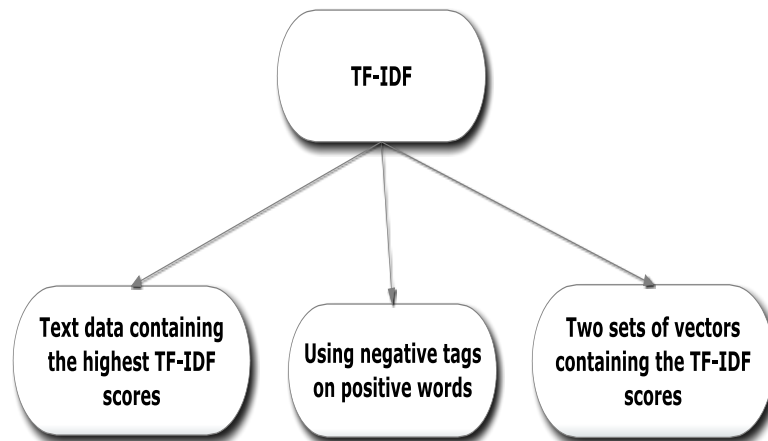
Figure 2: Flow of TF-IDF

FILTER STOP WORDS

► The uploaded data was then processed in the second component ('Process Documents to Data') using several subcomponents ('Extract Content', 'Tokenize', 'Transform Cases', 'Filter Stop-words', 'Filter Tokens,' respectively) that filtered excess noise (misspelled words, common stop words, etc.) to ensure a uniform set of variables that can be measured.
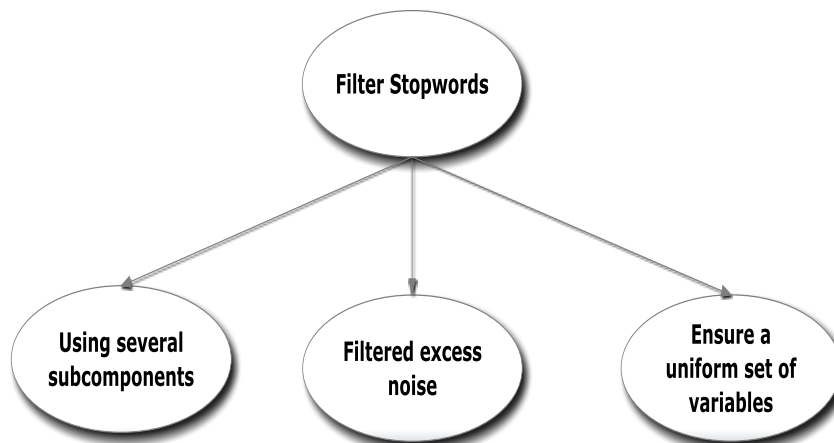
Figure 3: Flow of Filter Stopwords

SIDE EFFECT

In step of our network-based examination, we planned a plan for finding possible side effects occurring during the treatment and which user posts on the forum highlight. To this goal, we overlap the TF-IDF notches of the additional word list on to modules. The TF-IDF notches within each unit will thus straight reflect how recurrent a certain side-effect is stated in module

SYSTEM ARCHITECTURE
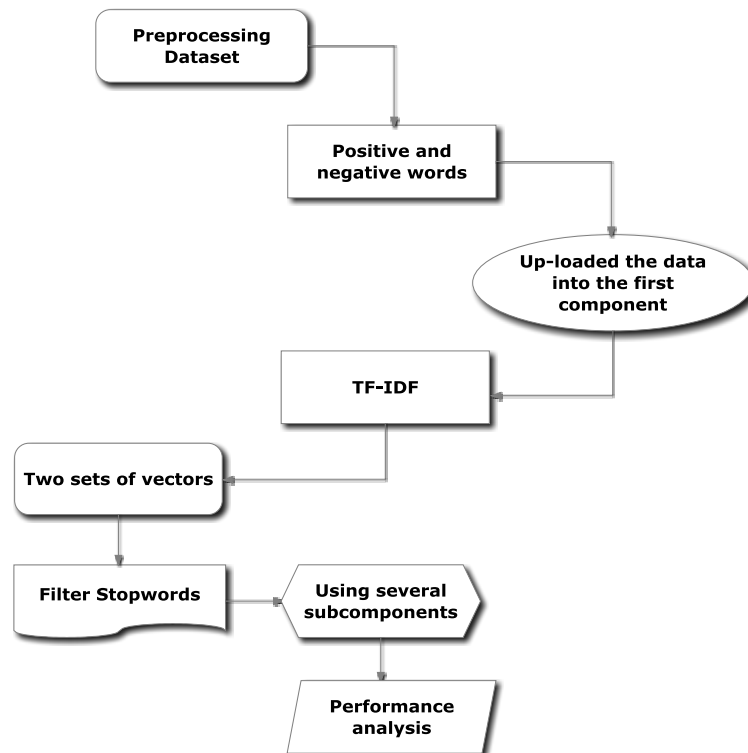
CONCLUSION

Social media is one of the emerging fields which provide a data on web for accessing the data through web. Recently it was used for the field of Health Informatics. Research done in Health Informatics to find the symptoms and side effects of the drugs based on the data mining approach. For this the system uses the approach called SOMs. Social media is one of the emerging fields which provide a data on web for accessing the data through web. Recently it was used for the field of Health Informatics. Research done in Health Informatics to find the symptoms and side effects of the drugs based on the data mining approach. For this we use the approach called SOMs and hierarchical clustering. This paper shows the framework which concentrates on positive and negative symptoms of the disease and also the side effects of the treatment in prevalent cancers lung cancer.

BIBLIOGRAPHY AND REFERENCES

[1] A. Ochoa, A. Hernandez, L. Cruz, J. Ponce, F. Montes, L. Li, and L.Janacek. "Artificial societies and social simulation using ant colony, par-ticle swarm optimization and cultural algorithms," New Achievements inEvolutionary Computation , P. Korosec, Ed. Rijeka, Croatia: InTech, pp.267–297, 2010.

[2] W. Cornell and W. Cornell. (2013). How Data Mining Drives Pharma:Information as a Raw Material and Product . [Online]. Available:http://acswebinars.org/big-data

[3] L. Toldo, "Text mining fundamentals for business analytics," presented atthe 11th Annual Text and Social Analytics Summit, Boston, MA, USA,2013.

[4] L. Dunbrack, "Pharma 2.0 – social media and pharmaceutical sales andmarketing," in Proc. Health Ind. Insights, 2010, p. 7.

[5] C. Corley, D. Cook, A. Mikler, and K. Singh, "Text and structural datamining of influenza mentions in web and social media," Int. J. Environ.Res. Public Health, vol. 7, pp. 596–615, Feb. 2010.

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