Link Based Overlapping Community Detection and Medical Data Mining Of Social Media for Cancer Prognosis

Ambika P, Binu Rajan M.R

Abstract*: Social media, ranging from personal messaging to live foras, is providing unlimited opportunities for patients to exchange their views on their experiences with drugs and devices. Here the aim is to understand the correlation between user posts and positive or negative judgment on drugs along with its side effects in cancer patients with particular emphasis on analysing the notion of community detection within this social network by analysing link properties. The proposed system is a two-step analysis framework where positive negative user sentiments are evaluated using data mining tools and techniques followed by identifying overlapping community structures(influential user modules) within the user forum. The two-way process utilizes the comments on internet message boards(cancer research forums) to infer the acceptance and effectiveness of a drug in cancer treatment and maps to the influential user within the network. In the first stage of the current study, opinion labels are developed about each drug based on opinion analysis from user posts and each word is given weightage per node using data mining tools. In the second stage, networks are built from the search results of the forum, a network ranking system reflecting the opinion formation about the drug is developed. Different from traditional algorithms based on node clustering, the proposed method is based on link clustering to discover overlapping communities. Since links usually represent unique relations among nodes, the link clustering will discover groups of links that have the same characteristics. The current approach effectively searches for different levels of organization within the networks and uncovers dense modules using partition density factor. Finally, the accuracy of novel link based overlapping community detection method is compared with the traditional network based community detection model using graph benchmark. Thus the experiment is used to determine opinion from consumer and identify influential users within the retrieved modules using information derived from both term occurrence and word frequency of data and network-based properties in an accurate way.

Index Terms*: Community detection, Health Informatics, Multi-scale, Markovprocess, Modularity, Overlapping communities, Random walks, Social media, Stability.

I. INTRODUCTION

With developing advanced web technologies social media has become a ubiquitous communication platform where users all around the world can share information related to their common interest. Hence online social network shares a common behaviour pattern or periodic interaction pattern. Thus by gaining an insight to these social norms of individual behaviour and human interaction and combing them with the reviews and computation of virtual community a methodical or an organized analysis could be done[1].Thus OSN sites provides limitless opportunities to pharma companies, for implementing viral marketing, target marketing[7] and for better business intelligence. Now a days patient rely mostly on social media to understand more about diseases, drugs and treatments as much of the information they receive from traditional sources are difficult to understand and interpret. Patients gets openness to experience with drugs and devices and health care providers can get feedbacks on their products and services and could use patient opinion, consumers' knowledge to improve their services. Also physicians could gain insight in improving the treatment recommendation from similar diagnosis methods from results obtained from these social forums. To gain inept knowledge of social media dynamics several data mining techniques such as graph prediction[3],opinion mining[4,5],sentiment analysis[6],link mining, link prediction [8,9]are utilized. Since the biomedical text mining[2]extracts data from unstructured, noisy environment researchers could make use of several advanced knowledge information retrieval systems to mine information in an accurate way.

Several patient oriented sites like cancerforum.net, WebMD, Medivizor, patientslikeme are available where the information shared among the users as feedback or user experience could be modelled using several network modelling techniques. Social networks can be represented as graphs for visual convenience and is a commonly used tool to model and analyse the structure of internal dynamics between a set of entities. Graphical representation of network with nodes and interconnecting links gives a visual way of its global organization [10]. Community detection within a network points out in finding densely connected groups or closely knitted communities of nodes. These nodes will be internally connected to the nearby nodes within the group than with the rest of the nodes within the network. Compared with the traditional methods like surveying, manual data collection, feedback forms with the help of social media and web crawling, scraping tools and willingness to share user generated information real time monitoring of OSN is made possible [11].

Real world entities within the network can interact with each other resulting in multiple scales of organisational hierarchies where modules often overlap with properties or functions of nodes of different community. To identify the structure of overlapping communities several algorithms have been proposed which can be broadly classified as node based and link based multi scale algorithms. Community detection algorithm mainly focuses on topological structure and layout

Revised Version Manuscript Received on October 07, 2016.

Ambika P, Department of Computer Science & Engineering, SCT College of Engineering, Trivandrum (Kerala)-695018 India.

Binu Rajan M.R, Department of Computer Science & Engineering, SCT College of Engineering, Trivandrum (Kerala)-695018 India.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

of the network while traditional clustering algorithms mostly consider node attributes. Node based algorithms make use of node similarities where as latter method is based on the intuition that a link in networks usually express the unique relation, placing each link in a single context reveals hierarchical and overlapping relationships of the communities detected. The novelty of proposed system lies in the fact that overlapping communities are determined using link properties compared with traditional node based modeling approaches. The information is rendered from internet message boards to infer the acceptance of drug in cancer treatment. The emphasis is on determining influential user in a much more effective way making use of inter weights and intra weights of links and partial density function as optimization factor. The correlation between labeled opinions obtained from user posts are mapped using several effective open data mining toolkits. Next stage is followed by constructing a network based on user posts and identifies user communities based on novel a overlapping detecting algorithm LBoCD. The accuracy of the cluster obtained is finally checked with truth value of any existing benchmark graph. Empirical experiments forum networks reveal that LBoCD can effectively detect the overlapping structure.

II. RELATED WORK

Altug Akay et al. proposed a novel data mining method[12] to monitor the experience of the drug Sitagliptin by diabetes mellitus type 2 patients. Using SOM the user opinion structure was analysed from forum posts followed by network modelling using depth first search method. The first step resulted in determining correlation between user clusters and user opinion. These findings results in new channels of research into prompt data collection, feedback, and analysis and the result improved health informatics and pharmaceutical manufactures.

Much similar work was put forward [13] community to improve healthcare outcomes and reduce costs using consumer-generated information by ascertaining user opinion for drug traceva. A network partitioning method based on optimizing a stability quality measure was employed. This allowed to determine consumer opinion and identify valid context posters within the community using information derived from user posts. Another investigatory analysis using the simulated SOMs[14] was used to check the correlations between user posts and positive or negative sentiments and opinion on drug. In [11] the hierarchical clustering, was used starting with an empty network of n vertices and no edges, in order of decreasing similarity one edge is added at a time between pairs of, starting with the pair with strongest similarity to find out strongly connected user clusters.

A genetic algorithm GaoCD[15] make use of the property of links for detecting overlapping communities. The link clustering algorithm considers an objective function partition density D for optimization and corresponding operators and genotype representation determines community automatically without any prior information.

Latent Dirichlet Allocation (LDA)-Based Link Partition (LBLP)method[16] finds communities with an adjustable range of overlapping by using link partitions. The algorithm calculates community belonging factor Interaction Profile(IP) for each link employ to encode edges and LDA to infer the probability that edges belong to communities. On the basis of this probability, link partitions with bridge links are determined in an efficient and faster way.

Le Martelot et al. proposed a stability optimisation [17,18] for determining the quality of partition obtained within a subgraph. The quality measure considered was over all Markov time as a resolution parameter in a given interval using greedy approaches. Modularity as optimisation criterion computes the change in modularity between initial partition and new partition where the clusters are merged.

III. PROPOSED MODEL

The proposed system is a two-step analysis framework where positive negative user judgements are analysed using data mining techniques and tools followed by identifying overlapping community structures(influential user modules) within the user forum. The two-way process utilizes the comments on internet message boards(cancer research forums) to infer the acceptance and effectiveness of a drug in cancer treatment and maps to the influential user within the network. In the first stage of the current study, opinion labels are developed about each drug based on opinion analysis from user posts and each word is given weightage per node using data mining tools. In the second stage, networks are modelled using link based overlapping community detection approach from the search results of the forum, which reflects the extend up to which each network is involved in the opinion formation about the drug. The frame work of entire work is as depicted.

A. Data Exploration

Several patient oriented sites like cancer.net, patientslikeme, BREASTCANCER.org, Smart Patients, Cancer Survivors Network, e-patients.net, Webmed etc. provides unlimited opportunities for patients to share their experiences with drugs and devices. We extracted consumer-friendly information, posts from most relevant cancerforum.net regarding most popular cancer treatment drugs, view ratings, user reviews and more. The commonly used oncology drugs such as xelox, brutinib, cisplatin, taxol were searched and resulted posts were compiled by rendering information from HTML pages.

B. Data Analysis & Pre-Processing

Text pre-processing is done using rapidminer toolkit which creates word vectors from the collection of posts. Various operations as tokenize, transforming cases, filtering, stopword elimination are used to process and the result summary gives document occurrence and word occurrence of each tokens. Here TFIDF is used to calculate occurrence of each word over the whole corpus document and the fraction of document made up of the word.

Weight for each processed word from user post is assigned as

$tf\text{-}idf_{t,d}=tf_{t,d} x idf_t$

df_t is the document frequency of t: the number of documents that contain t

$idf_t= \log N/df_t$

N is NumberOfDocuments

Only words with high weightage are used for further analysis.

C. Opinion Categorising and Sentiment Statistics

Each word in the post is marked to a particular

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

part-of-speech using Apache Open NLP library which is a machine learning based toolkit for the processing of natural language text. It make use of Maximum Entropy Model (MaxEnt) for POS tagging (part-os-speech tagging) NLP task. The words with selected POS tags are included from the corpus. The processed words are fed to a lexical resource Senti Word Net which assigns positivity, negativity, objectivity[19].The words with highest TFIDF score is choose and the resulting lists of positive and negative words are formulated from each post. Based on the senti word net score each sentence is marked as strong positive, weak positive, positive, strong negative, negative, weak negative. Depending upon polarity of each sentence, each words are listed in corresponding word list. The insignificant outliers are eliminated from the word list by including words that appeared less than a threshold value.

The MeSH (National Library of Medicine's Medical Subject Heading) vocabulary is searched for each words in the post to obtain the side effects (http://www.nlm.nih.gov/mesh/). The words annotated with specific qualifiers CI – chemically induced; CO – complications; DI – diagnosis; PA – pathology, and PP – physiopatholog are selected. The words with highest TFIDF score will be included in the sideffect lists.

The correlations between user posts and positive or negative judgment on drugs are visually analysed using the self-organizing maps (SOMs). The SOM toolbox[14]was used for analysis and the positive and negative clusters are formulate form the first vector word list in MATLAB. The weight values reflect on the cluster content.

D. Network Modelling

The user posts are extracted from the sites and the links between each user(nodes) are determined. Each link is assigned with the user postsize based on the reply given to each posts. The thread initiators will start a conversation and the context posters reply to each other. From the internal dynamics of the network a directional graph is constructed depending upon the flow of information. Each thread is identified with individual forum id and may have hundreds of users with several user overlapping with other context thread. Here the aim is to determine only the influential user who is capable of redirecting the information flow within the network.

E. Community Detection Algorithm

Network as a whole is considered and the minimum edge is removed from the cluster. The edge weight is formulated as

TotalWeight=posts.size+ γ*InterWeight; where the edge weight is governed by the parameter γ. In overlapping communities, one node can belong to more than one community, which makes the conventional community definitions unreasonable. Hence the InterWeight factor determines whether connected node belongs to same subgraph or different. The algorithm further optimises the subgraph identified based on the partial density factor which determines the fitness of each node.

For a network with M links, suppose P = {P1, ⋯,P_C} as a partition of the links into C subsets. $m_c = |P_c|$ is the number of links in subset c.

The partition density factor D is the average of Dc over all communities, weighted by the fraction of links presenting in each community. Dc represents the link density of each community.

$D=\sum_c (m_c/M)D_c$

Different from the conventional community evaluation criteria that a community should be densely intra-connected and sparsely connected with the rest communities, partition density D evaluates the link density within each community, which is suitable for overlapping community detection. Density D is calculated at each step removing the minimum edge and is used to determine the best node resulting in highest density subgraph or cluster in each iteration. No of clusters is assumed to a threshold value node.size/2.Each cluster will be having nodes with highest participation.

F. Algorithmic Implementation

Algorithm 1: Link based overlapping community detection Algorithm : LBoCD

Input: Given data set, Threshold γ
Output: Best found partition C

1. Create the network with as many nodes as there are in the user post
2. Set this partition as current partition and as best known partition C
3. Compute partition density vector D
4. Determine minEdge by computing the total weight as
5. TotalWeight=posts.size+ γ InterWeight;
6. Remove the minimum edge and store it in removed node list
7. Compute the partial density factor for each resulting subgraph as density_selected
8. **while** 2 subgraphs at least are possible in current partition **do**
9. **forall** edges from the removed list **do**
10. **If** new density>orginal undivided graph density **then**
11. Keep in memory new partion of communities replace C with the best partion
12. density_selected= D
13. **end if**
14. **end for**
15. **end while**
16. Check for all other remaining edges
17. Return best found partition C

G. Side effect Analysis And Total Satisfaction

The community detected through LBoCD is further refined from the vectorlist obtained through data mining tool as well as the side effects filtered from mesh vocabulary. The post specifying side effects is given more weightage and the users which spread the opinion among other users are selected based on average threshold value obtained from word frequency. Those users are marked as selected users within the information module. The global variable considers total post count as the posts can be positive negative or neutral.

$Value_i$ = (Sum_+-Sum_--SideeffectValue)/Total.Posts

Sum_+ represents the total sum of the TF-IDF scores matching the positive words in the wordlist vectors within the module. Similarly $Sum_-$is the total sum of the TF-IDF scores matching the negative words in the wordlist vectors within the module.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

The selected nodes within the modules will be having highest degree.

Also as in Altug Akay *et al*. proposed two variables module average opinion and user average opinion to determine TF-IDF score matching the nodes in a specific module.

$$MAO = \frac{Sum_{+} - Sum_{-}}{Sum_{all}}$$

However the method offers a limitation in determining the influential user. As if a particular user has submitted 100 positive posts and two negative posts the MAO results in a value less than one. But if a user has posted only 5 positive posts as per former method the weightage is given to second user which results in converging to a false negative value. Hence the proposed method works much more accurately in determining the information broker. Similarly the total satisfied and dissatisfied user counts per module are determined from the scores that reflects the consistency of positive and negative opinion.

IV. RESULTS & DISCUSSION

A. *Experimental Evaluation*

Experiments on user forum includes (1) The effectiveness of Link based on overlapping community detection is evaluated on user posts (2)The structural characteristics of communities discovered by LBoCD with the ground truth value of any benchmark community detection algorithm. 3) Comparison of traditional MSCD method to proposed system.

Community detection divides a large network into groups of nodes, where nodes are densely connected within but sparsely connected outside. Modularity was initially introduced to evaluate quality of partitions latter broadened to optimisation function. Modularity optimization faces from several limitations. Optimisation methods using modularity to determine the quality of partition within a network, can fail to detect small communities or over-partition networks. Another concern is that lacks a clear global maximum value. Modularity is locally optimised at each step of partition assuming the current partition to be the best one. Stability optimisation tends to settle for longer on fewer partitions than modularity approach. As internal connection between communities becomes denser with increasing number of overlapping nodes modularity function modularity Q does not fit for overlapping communities. In [13], the authors proposed an approach in which transition probabilities for a random walk of length t (t being the Markov time) enable multiscale analysis. With increasing scale t, larger and larger modules are found where At is the adjacency matrix, t is the length of the network, m is the number of edges, i and j are nodes, di is node i's (and j's) strength, and S (i,j) function becomes one if one of the nodes belong to the same network and zero if it does not belong to any network.

$$Q_t = \sum_{i,j} \left(At - \frac{didj}{2m} \right) * S(i,j)$$

The proposed method make use of a objective function partial density factor much similar to the one used in [15,21]. Most of the node based algorithms need prior information to detect overlapping multi scale communities and usually they could not provide the global topological structure of networks. The link-based method emphasizes on the unique role that each link represents, and provides a new way for overlapping community detection. Different from the traditional community evaluation criteria the partition density D evaluates the link density within the community, which is suitable for overlapping community detection. The experiment for community detection was performed on varying size real world network modelled from the drug review from cancer forum.

Table 1. Real network modelled from user posts used in experiment

	Opdivo	Xelox	Ibrutinib	Taxol	Cisplatin
# of nodes	10	15	27	36	66
# of edges	17	40	91	121	178

B. *Analyzing Ground-Truth Communities*

The quality of the communities detected by the LBoCD algorithm on networks for which the true grouping is obtained from a benchmark algorithm has been verified. To study the behavior of the developed algorithm, extensive experiments have been conducted in real-world networks with varying node size and edge distribution. The quality of the resulting clustering is assessed by means of the similarity of cluster, computed between a given output clustering and the true one. First, for each testing graph another graph generated with the same settings as the dynamic data extracted from HTML pages are used. We use the partition density D [15] as the evaluation criterion, which evaluates the link density inside the communities. To test the performance of LBoCD quantitatively, we adopt k-Center Approximation [20] as benchmark and use similarity of cluster as measure criterion. If the road network of the city is modelled as an undirected graph whose edge weights are the distances between intersections, then this is an instance of the k-center problem. Similarly for distributing k services to a city with all service being accessed within a minimum reach falls under the category of k-centre approximation. In the k-center problem an integr k and a graph G = (V,E) with nonnegative edge weights are considered. The problem is to compute a subset of k vertices(centres) C _V , such that the maximum distance between any vertex in V and its nearest center in C is the minimum.

Through the experiment it has been found that LBoCD always has best performance when compared to traditional community detection algorithm which make use of node attributes and modularity as optimization parameter. Experiments showed that the optimization yields significant gain in speed and performance without any loss in accuracy. Fig 1 shows the performance comparison of two methods.

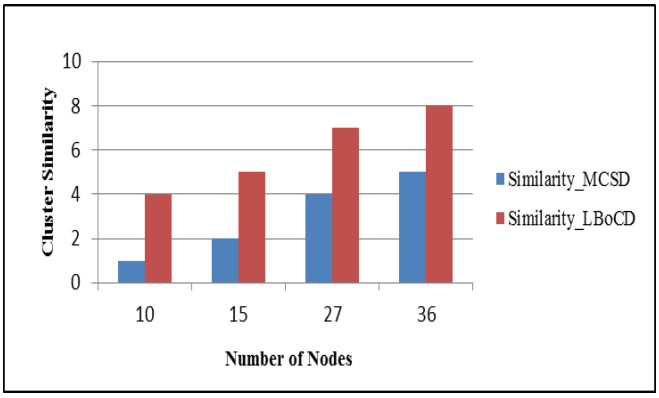

Fig 1:Performance comparison of LBoCD and MCSD on

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

the distribution of community size on real networks.
Fig 2 provides running time comparison for each method over several networks of varying node size and edges.

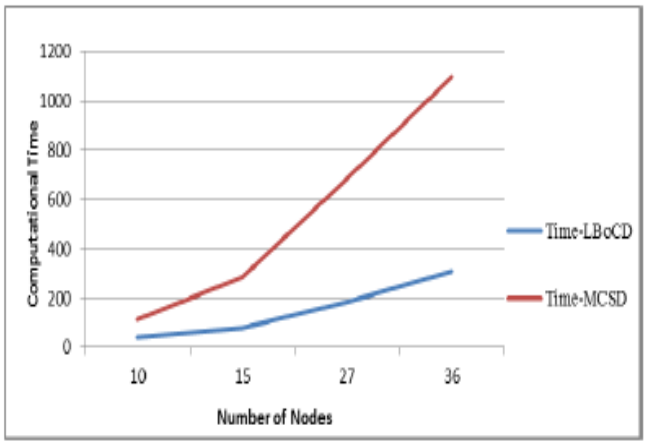

Fig 2: Running Time Comparison of LBoCD with variation of network size

C. Observation & Result

The experiments were conducted on the real word drug review dataset which is collected from http://www.cancerforums.net.Different oncology drugs such as xelox, ibrutinib, cisplatin, taxol are searched and threads from user posts are extracted from the forum. From the experiments, the system detects the best drug finding based on text reviews and ratings. The approach make use of a novel method of community detection and make use of the result to map to the sentiment analysis result to obtain the total satisfaction and dissatisfaction level.

A picture begins to emerge of the user's opinion that is roughly divided with regards to satisfaction of several oncology drugs. Another factor governing the negative opinion in judgement is side effect which is given additional support in each posts. Positive opinion grounds from long term experience and usage of drugs and influenced by similar comments. The major emphasis was given to community detection. Here the users can respond in several threads and hence determination of overlapping nodes is of prior concern. The entire network was considered for analysis which contains both connected and disconnected units. Further optimization was carried out using partial density factor and in each network The Densities of the retrieved modules range from 0.2 to 0.8. Further scrutinizing influential users within the modules revealed that they were informative and actively interacting with users across many threads.

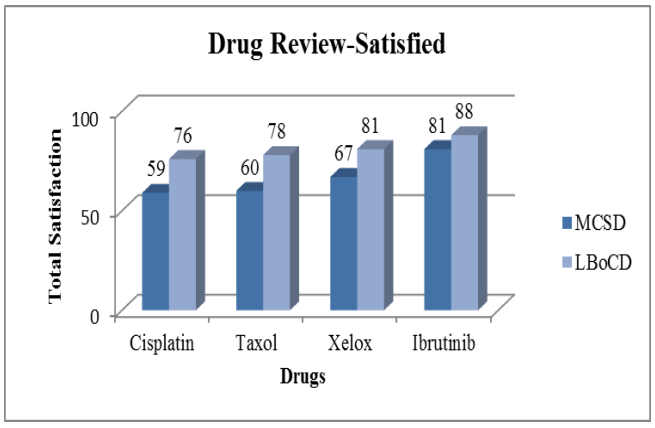

Fig 3: Drug Comparison Chart I

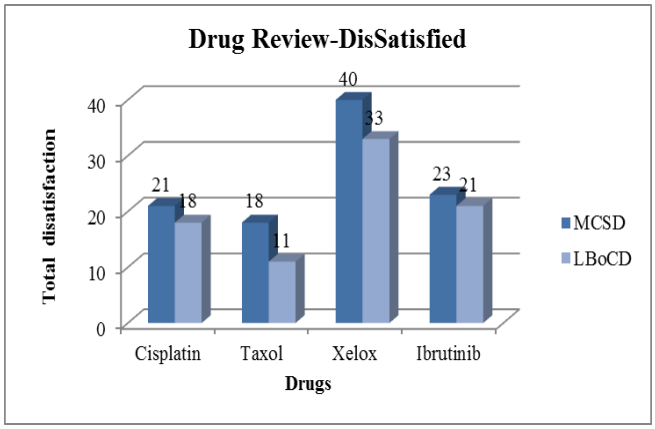

Fig 4: Drug Comparison Chart II

Further analysis was conducted on which specific side effects are covered under selected community.

The side effects extracted from MESH vocabulary were enriched with TFIDF scores and weightage is given per hieghest score.

Table 2:Side effects frequency on selected modules for Xelox

Module	Side effects	Value
Module 0	tumor	7
Module 0	rash	15
Module 0	stroke	5
Module1	ulcer	5
Module1	carcinoma	4
Module2	headache	13
Module2	dizziness	4

The experimental study reveals the result from LBoCD is more accurate compared to MCSD method using node attributes as constraints with and the precisions of all traditional algorithms is relatively low in detecting communities and overlaying those modules with polarity rating.

V. CONCLUSION

The paper depicts a novel link based community detection algorithm LBoCD and a framework which make use of the algorithm in pharmaceutical studies which determines positive negative opinion of oncology drugs and its side effects. The application modelling make use of underlying information among users in social media. Modules of strongly interacting users were identified using LBoCD algorithm. The identified modules are overlaid from the content information in the form of term occurrence and its frequency count of words retrieved from user posts and the influential users are determined. Additionally, potential side effects consistently discussed by community of users were identified accurately. Such an approach would pay way for monitoring, diagnosis, future prediction, prescribing similar treatment histories, disease prognosis and covers other related treatment issues. The approach aims at determining the effective drug in cancer diagnosis and prognosis, intelligent use of social media for pharmaceutical marketing and analyse web for user influence and opinion, for future treatments and to provide rapid,

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.

up-to-date information for the pharmaceutical industry, hospitals, and medical staff. Furthermore the application could be enhanced by considering rankings, 'likes' of posts, and friendships from user forums. Use of medical lexical dictionaries, Biomedical named entity recognition technique[23] for automatically identifying biomedical terms can be used to enhance the prediction and classification techniques.

Social media mining aims at providing a new arena where the big data could be analysed and process for knowledge extraction that enables support for cost reduction, decision making, target marketing, viral marketing and sharing health-related experiences.

REFERENCES

1. Reza Zafarani, Mohammad Ali Abbasi,Huan Liu, "Social Media Mining An Introduction,", April 2014.J. Cambridge university.
2. Zhu,F., Patumcharoenpol, P., Zhang, C., Yang, Y., Chan, J., Meechai, A. et al, "Biomedical text mining and its applications in cancer research. "J. Biomed. Inform. 2013;46:200–211.
3. David F. Nettleton,"Data mining of social networks represented as graphs," Expert Systems with Application, October 2012
4. G. Angulakshmi, Dr.R. Manicka Chezian,"An Analysis on Opinion Mining: Techniques and Tools",Intrn. Journal of Advanced Research in Computer and Communication Engineering Vol. 3,Issue 7, July 2014
5. Richa Sharma, Shweta Nigam and Rekha Jain,"Oinion mining of movie reviews at document level", International Journal on Information Theory (IJIT), Vol.3, No.3, July 2014.
6. WalaaM edhat a, Ahmed Hassan b, HodaKorashy," Sentiment analysis algorithms and applications: A survey", Ain Shams Engineering Journal (2014) ,Volume:5,Issue:4,pp: 1093–1113
7. Vishal Shrivastava, Rajesh Boghey, BhupendraVerma,"A Framework for Improving Target Marketing Using Collaborative Data Mining Approach", IJICT Journal, Volume 1 No. 2, June 2011
8. LiseGetoor," Link Mining: A New Data Mining Challenge,"UMIACS, 415- 444,Volume 4, Issue 2,2013
9. Mohammad Al Hasan , Mohammed J. Zaki,"A Survey of Link Prediction in Social Networks",Springer,March,2011
10. P.Ambika, M.R BinuRajan, "Multi-scale Community Detection in Complex Networks," IEEE International Conference on research Advances in Integrated Navigation System,2016.
11. V.R. Nagarajan, Monisha.P.M" Extracting Knowledge from Social Media toImprove Health Informatics"IJARCC,Vol. 4, Issue 7, July 2015.
12. 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. Akay, A. Dragomir, and B. E. Erlandsson, "Network-Based Modeling and Intelligent DataMining of Social Media for Improving CareVol:19, 2015
14. J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parhankangas, "Self-Organizing Map in MATLAB: The SOM Toolbox," in Proc. Matlab DSP Conf., Espoo, Finland, 1999, pp. 35–40.
15. Chuan Shi, Yanan Cai, Di Fu, Yuxiao Dong, Bin Wu, "A link clustering based overlapping community detection algorithm," Data & Knowledge Engineering, Elsevier, vol. 87, pp. 394–404, May 2013.
16. Le Yu,BinWu,Bai Wang,"LBLP :Link-Clustering-Based Approach for Overlapping Community Detection," ISSN, Volume 18, pp387-397, Number 4, August 2013.
17. Erwan Le Martelot , Chris Hankin, "Multi-scale community detection using stability optimisation," International Journal of Web Based Communities, v.9 n.3, p.323-348, June 2013
18. 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.
19. Esuli ,A., Sebastian,F.," SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining, "In: Proceedings of 3rd Conf. on Intrn. Language Resource and Evaluation,pp.417-422(2006)
20. Design and Analysis of Computer Algorithms1,David M. Mount, CMSC 451
21. Y.Y. Ahn, J.P. Bagrow, S. Lehmann, Link communities reveal multi-scale complexity in networks, Nature 466 (2010) 761–764.
22. Fei Zhu, Preecha Patumcharoenpol, Cheng Zhanga, Yang Yang b, Jonathan Chan ,Asawin Meechai , Wanwipa Vongsangnak , Bairong Shen ,"Biomedical text mining and its applications in cancer research ," Journal of Biomedical Informatics 46 (2013) 200–211

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication Pvt. Ltd.