

Neural network based web mining of social media for health informatics

Saranya R

Dept of computer science&Engg
Maha barathi Engineering College
China salem 606201,Tamil nadu,India
saranpoongodi28@gmail.com

Suganthi T

Dept of computer science&Engg
Maha Barathi Engineering College
China salem 606201,Tamil nadu,India

Abstract— Inteligently extracting knowledge from social media has recently attracted great interest from the Biomedical and Health Informatics community to improve healthcare simultaneously, outcomes and reduce costs using consumer-generated opinion. We propose a two-step analysis framework that focuses on positive and negative sentiment, as well as the side effects of treatment, in users' forum posts, and identifies user communities (modules) and influential users for the purpose of ascertaining user opinion of cancer treatment. We used a self-organizing map to analyze word frequency data derived from users' forum posts. We then introduced a novel network-based approach for modeling users' forum interactions and employed a network partitioning method based on optimizing a stability quality measure. This allowed us to determine consumer opinion and identify influential users within the retrieved modules using information derived from both word- frequency data and network-based properties. Our approach can expand research into intelligently mining social media data for consumer opinion of various treatments to provide rapid, up-to-date information for the pharmaceutical industry, hospitals, and medical staff, on the effectiveness (or ineffectiveness) of future treatments.

Keywords—Data mining, neural networks, complex networks, social computing, semantic web.

I INTRODUCTION

Social media is one on the extraction of information from the Internet. Now a day is used to extract the data from patients who know the experiences of the patient's symptoms. Social Media, offers immense possibilities for patients to maintain their experiences with drugs and devices. Social media allows message, post, information gathering and distribution in the healthcare space. It provides an effective social networking environment. Using network-based analysis method is to model the social media like Face book, Twitter, WebMD, etc. These analyzes are of nodes that can be individual or organization who join the network in various associations such interests, friendship, kinship, etc. Therefore, the presentation of information would be a graphical, made more convenient for the user visualization.

Network especially when shared with web crawling and scraping software allows real-time monitoring of changes within the network. In social networks, it provides an in-depth knowledge of the social network dynamics. It used for replication studies of different network aspects such as mining.

The users broadcast knowledge among themselves Another example is the analysis of the proliferation of certain edges of the networks and how certain information is included develop the extensions. In social media the obligation of social networks includes data collection difficult way. It is used several techniques to employees, such as links, existence, estimation, object, group and subgroup discovery and mining the data, link mining, classification by links, predictions based on objects. Basis were they link- prediction, viral marketing, online discussion groups (and rankings) enable the development of solution.

II. RELATED WORK

The community detection received an increasing attention as a way to uncover the formation of networks by grouping nodes into communities more densely connected internally than superficially[1] Several heuristics are commenced for speed-up purposes. Experiments exhibit the efficiency and exact of our method with respect to each algorithm and criterion by testing them against large generated multi-scale networks [2].

Time series and cross sectional aggregation of message information progress the superiority of the resultant sentiment directory.

Frequent sub graph mining is an active research topic in the data mining community. A graph is a common form to signify data and has been used in many domains like informatics and bioinformatics[3]. A sensor located inside a digital camera is only able to measure the light which is reflected by an object. The reflected light varies with the spectral power distribution of the illuminant. Hence, images taken with a digital camera may show a strong color cast if an incorrect white balance setting has been chosen. Such a color cast may also be due to an automatic white balance not working correct. In contrast, colors perceived by a human observer appear to be approximately constant.

The purpose is to arrive at recognition of multicolored objects invariant to a substantial change in viewpoint, object geometry and illumination. Assuming dichromatic reflectance and white illumination, it is shown that normalized color rgb , saturation and hue, and the newly proposed color models are all invariant to a change in viewing direction, object geometry and illumination[5]. a comparative analysis of algorithm performance that we use as the basis of a discussion of the current state of colour constancy research and of the important issues that future research in this field should address. Finally, some areas of recent research are highlighted that are important in the context of further improving the performance of colour constancy algorithms[6].

Color constancy is the skill by which it is possible to tell the color of an object even under a colored light[7]. I interpret the color of an object as its color under a fixed canonical light, rather than as a surface reflectance function.

Christo Ananth et al. [8] proposed a system which is an innovative congestion control algorithm named FAQ-MAST TCP (Fast Active Queue Management Stability Transmission Control Protocol) is aimed for high-speed long-latency networks. Four major difficulties in FAQ-MAST TCP are highlighted at both packet and flow levels. The architecture and characterization of equilibrium and stability properties of FAQ-MAST TCP are discussed. Experimental results are presented comparing the first Linux prototype with TCP Reno, HSTCP, and STCP in terms of throughput, fairness, stability, and responsiveness. FAQ-MAST TCP aims to rapidly stabilize high-speed long-latency networks into steady, efficient and fair operating points, in dynamic sharing environments, and the preliminary results are produced as output of our project. The Proposed architecture is explained with the help of an existing real-time example as to explain why FAQ-MAST TCP download is chosen rather than FTP download.

III PROPOSED SYSTEM:

In the novel system study, the exploratory analysis using the self-organizing map(SOMs) to assess correlations between user posts and positive or negative opinion on the drug. In the second stage, we model the users and their posts using a network-based approach. We build on our previous study and use an enhanced method for identifying user communities (modules) and influential users therein. This approach effectively searches for potential levels of organization (scales) within the networks and uncovers dense modules uploaded the data into the first component (‘Read Excel’). 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. The final component (‘Processed Data’) contained the final word list, with each word containing a specific TF-IDF score.

This leads to an analysis that shows two distinct sets of circumstances under which color constancy is possible. In this framework, color constancy requires estimating the illuminant under which the image was taken. The estimate is then used to choose one of a set of linear maps, which is applied to the image to yield a color descriptor at each point. This set of maps is computed in advance..

By overlaying these modules with content-based information in the form of word-frequency scores retrieved from user posts, we were able to identify information brokers which seem to play important roles in the shaping the information content of the forum

I interpret the color of an object as its color under a fixed canonical light, rather than as a surface reflectance function. This leads to an analysis that shows two distinct sets of circumstances under which color constancy is possible. In this framework, color constancy requires estimating the illuminant under which the image was taken

IV.Simulation and Conclusion

Digital computers are used to process the image. The image will be converted to digital form using a scanner – digitizer and then process it. It is defined as the subjecting numerical representations of objects to a series of operations in order to obtain a desired result. It starts with one image and produces a modified version of the same. It is therefore a process that takes an image into another.

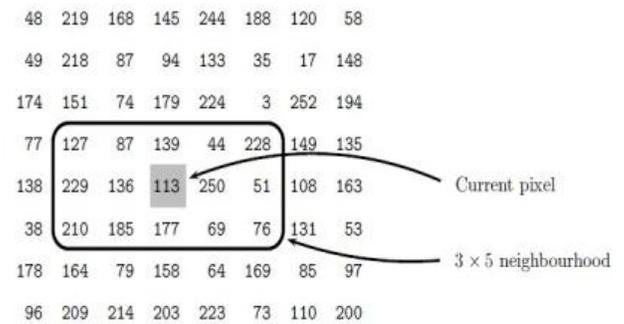
The term digital image processing generally refers to processing of a two-dimensional picture by a digital computer. In a broader context, it implies digital processing of any two-dimensional data. A digital image is an array of real numbers represented by a finite number of bits

The principle advantage of Digital Image Processing methods is its versatility, repeatability and the preservation of original data precision.

The various Image Processing techniques are:

- Image representation
- Image preprocessing
- Image enhancement
- Image restoration

- Image classification
- Image reconstruction
- Image data compression

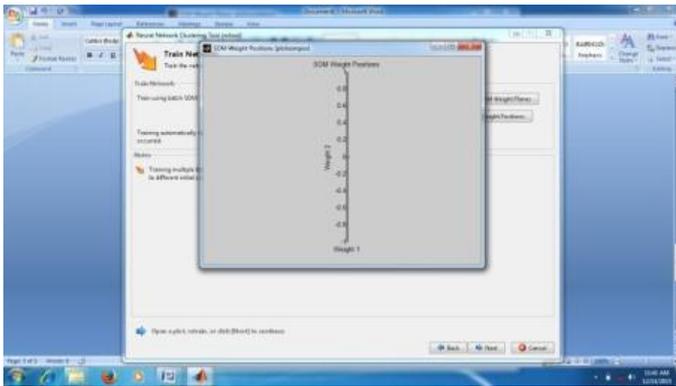
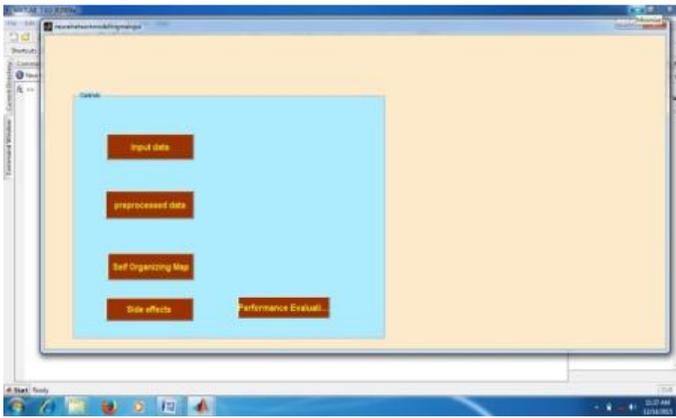


We converted a forum focused on oncology into weighted vectors to measure consumer thoughts on the drug Erlotinib using positive and negative terms alongside another list containing the side effects. Our methods were able to investigate positive and negative sentiment on lung cancer treatment using the drug by mapping the large dimensional data onto a lower dimensional space using the SOM.

In this solution will require more advanced detection of inter social dynamics and its effects on the members : such interests of study may be include rankings, 'likes' of posts, and friendship .The context posting will require formal languages dictionaries that include medical terms for specific diseases ,and informal languages terms('slang') to the clarify posts.The different platform will allow up-to-date information on the status of the drug in case one social platform ceases to discuss the drug. Another solution can look at multiple wordlists that can be include multiple treatments that, when combined with contextual posting and medical lexical dictionaries, can point to the source

Social media can open the door for the health care sector in address cost reduction ,product and service optimization and patient care.

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.



Suppose we take an image, a photo, say. For the moment, let's make things easy and suppose the photo is black and white (that is, lots of shades of grey), so no colour. We may consider this image as being a two dimensional function, where the function values give the brightness of the image at any given point. We may assume that in such an image brightness values can be any real numbers in the range (black) (white).

A digital image from a photo in that the values are all discrete. Usually they take on only integer values. The brightness values also ranging from 0 (black) to 255 (white). A digital image can be considered as a large array of discrete dots, each of which has a brightness associated with it.

REFERNCES

[1] A. Ochoa, A. Hernandez, L. Cruz, J. Ponce, F. Montes, L. Li, and L. Janacek. -Artificial societies and social simulation using ant colony, particle swarm optimization and cultural algorithms,| *New Achievements in Evolutionary 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 at the 11th Annual Text and Social Analytics Summit, Boston, MA, USA, 2013.

[4] L. Dunbrack, -Pharma 2.0 – social media and pharmaceutical sales and marketing,| in *Proc. Health Ind. Insights*, 2010, p. 7.

[5] C. Corley, D. Cook, A. Mikler, and K. Singh, -Text and structural data mining 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.

[7] Q. And and L. Getoor, —Link-based classification,| in *Proc. 20th Int. Conf. Mach. Learning*, Washington, D.C., USA, 2003, pp. 496–503Lu..

[8]Christo Ananth, S.Esakki Rajavel, I.AnnaDurai, A.Mydeen@SyedAli, C.Sudalai@UtchiMahali, M.Ruban Kingston, “FAQ-MAST TCP for Secure Download”, *International Journal of Communication and Computer Technologies (IJCCTS)*, Volume 02 – No.13 Issue: 01 , Mar 2014, pp 78-85

[9] B. Taskar, M. Wong, P. Abbeel, and D. Koller, -Link prediction in relational data,| in *Proc. Adv. Neural Inform. Process. Syst.*, Vancouver, B.C. Canada, 2003.

[10] D. Liben-Nowell and J. M. Kleinberg, -The link prediction problem for social networks,| *J. Am. Soc. Inform. Sci. Technol.*, vol. 57, pp. 556–559, May 2007.