ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

Improving Health Care Based on opinion Mining Using KNN

T.Anisha[1], Mr.N.Thulasi[2]

PG Scholar, Dept. of CSE, Dhanalakshmi Srinivasan College of Engineering & Technology, Chennai, India [1]

Associate Professor, Dept. of CSE, Dhanalakshmi Srinivasan College of Engineering & Technology, Chennai, India [2]

ABSTRACT: Intelligently extracting knowledge from social media has recently attracted great interest from the Biomedical and Health Informatics community to simultaneously improve healthcare outcomes and reduce costs using consumer-generated opinion. this 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. 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. This 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: HMM, CRF,

I. INTRODUCTION

Sentiment analysis is a type of natural language processing for tracking the mood of the public about a particular product or topic. Sentiment analysis, which is also called opinion mining, involves in building a system to collect and examine opinions about the product made in blog posts, comments, reviews or tweets. Sentiment analysis can be useful in several ways. For example, in marketing it helps injudging the success of an ad campaign or new product launch, determine which versions of a product or service are popular and even identify which demographics like or dislike particular features.

There are several challenges in Sentiment analysis. The first is an opinion word that is considered to be positive in one situation may be considered negative in another situation. A second challenge is that people don't always express opinions in a same way. Most traditional text processing relies on the fact that small differences between two pieces of text don't change the meaning very much. In Sentiment analysis, however, "the picture was great" is very different from "the picture was not great". People can be contradictory in their statements.

Most reviews will have both positive and negative comments, which is somewhat manageable by analyzing sentences one at a time. However, in the more informal medium like twitter or blogs, the more likely people are to combine different opinions in the same sentence which is easy for a human to understand, but more difficult for a computer to parse. Sometimes even other people have difficulty understanding what someone thought based on a short piece of text because it lacks context.

Classifying evaluative texts at the document level or the sentence level does not tell what the opinion holder likes and dislikes. A positive document on an object does not mean that the opinion holder has positive opinions on all aspects or features of the object. Likewise, a negative document does not mean that the opinion holder dislikes everything about the object. In an evaluative document (e.g., a product review), the opinion holder typically writes both positive and negative aspects of the object, although the general sentiment on the object may be positive or negative.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

To obtain such detailed aspects, going to the feature level is needed. Based on the model presented earlier, three key mining tasks are:

1. Identifying object features: For instance, in the sentence "The picture quality of this camera is amazing," the object feature is "picture quality". In a supervised pattern mining method is proposed. In an unsupervised method is used. The technique basically finds frequent nouns and noun phrases as features, which are usually genuine features. Clearly, many information extraction techniques are also applicable, e.g., conditional random fields (CRF), hidden Markov models (HMM), and many others.

2. Determining opinion orientations: This task determines whether the opinions on the features are positive, negative or neutral. In the above sentence, the opinion on the feature "picture quality" is positive. Again, many approaches are possible. A lexicon-based approach has been shown to perform quite well. The lexicon-based approach basically uses opinion words and phrases in a sentence to determine the orientation of an opinion on a feature. Clearly, various types of supervised learning are possible approaches as well. 3. Grouping synonyms: As the same object features can be expressed with different words or phrases, this task groups those synonyms together. Not much research has been done on this topic. An attempt on this problem.

An important part of information-gathering behavior has always been to find out what other people think. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges arise as people can, and do, actively use information technologies to seek out and understand the opinions of others. The sudden eruption of activity in the area of opinion mining and sentiment analysis, which deals with the computational treatment of opinion, sentiment, and subjectivity in text, has thus occurred at least in part as a direct response to the surge of interest in new systems that deal directly with opinions as a first-class object.

This survey covers techniques and approaches that promise to directly enable opinion-oriented information-seeking systems. This focus is on methods that seek to address the new challenges raised by sentiment-aware applications, as compared to those that are already present in more traditional fact-based analysis. It include material on summarization of evaluative text and on broader issues regarding privacy, vulnerability to manipulation, and economic impact that the development of opinion-oriented information-access services gives rise to. To facilitate future work, a discussion of available resources, benchmark datasets, and evaluation campaigns is also provided.

II. OBJECTIVE OF THE PROJECT

To effectively handle valuable user feedback data(ontology) to evaluate drugs based on user polarity.To detect inter-social dynamics and its impacts on other members.To accumulate and process user based data from several forums and discussion groups simultaneously.To facilitate generation of an Independent algorithm to handle medical instruments also.It focusses on converting the ontology based data to weighted data.It helps in advanced identification of the inter-social dynamics like ranking.Further emphasizes on inclusion of formal language and medical lexical dictionaries for most prevalent diseases.More precisely pinpoints the reasons for user satisfaction/dissatisfaction on the item, here drug.Use of enhanced mathematical techniques.Inter-social dynamics helps in discovering the relationship between user opinions.Inclusion of medical lexical dictionaries helps in analyzing for most other diseases.Gives a clear picture on the sources of user satisfaction/dissatisfaction.

III. PROPOSED DESIGN

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

➢ Methods for preparing input validations and steps to follow when error occur.

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

The output form of an information system should accomplish one or more of the following objectives.

- ❖ Convey information about past activities, current status or projections of the
- ❖ Future.
- ❖ Signal important events, opportunities, problems, or warnings.
- ❖ Trigger an action.
- ❖ Confirm an action.

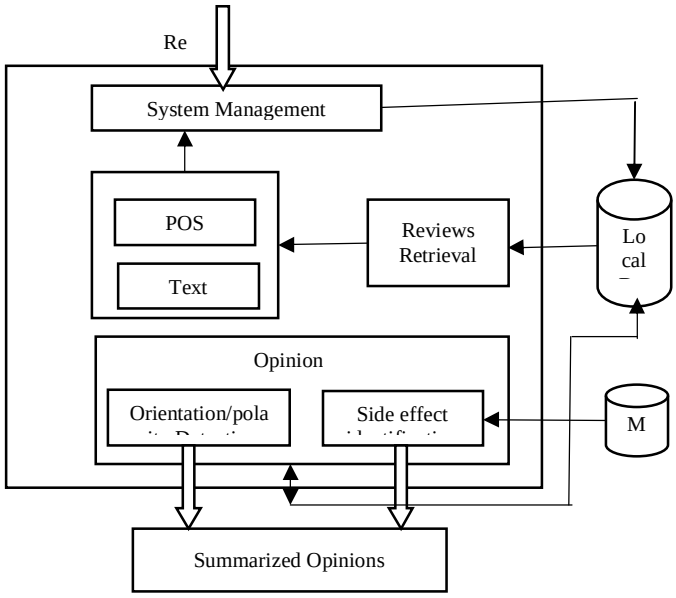

Fig. 1 System Specification

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

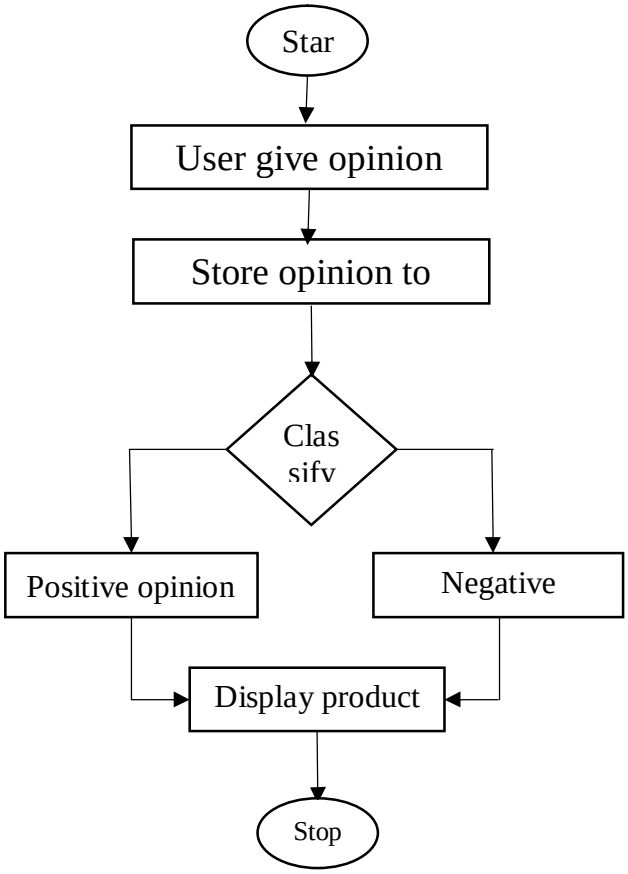

Fig. Flow Chart Diagram

The data pre-processing involves collecting opinionated text documents, performing initial data cleaning and transforming relevant text data into the input variables form as needed by self-organizing maps. In this study to have performed tokenization, stemming and stop word removal in data pre-processing. It have used Vector Space Model (VSM) to generate the bag of words representation for each document. In VSM, the weights represent the tf·idf score which is computed from term frequency (tf) and inverse document frequency (idf). The tf·idf weight for term j in document Di is computed as log (1) Where fij is frequency of term j in document I and dfj is number of document having term j.

The self-organizing map (SOM) is an unsupervised competitive ANN based on the winner-takes-all (WTA) principle (Kohonen, 1995). A typical SOM is made up of a vector of nodes for input (input neurons layer), an array of nodes as output map, and a matrix of connections between each output layer unit and all the input layer units. The input nodes receive vectorized input data patterns and propagate them to a set of output nodes, which are arranged according to map topology (generally two-dimensional grid) forming so called Kohonen's layer. Kohonen's principle of topographic map formation, determines how the spatial location of an output node in the topographic map corresponds to a particular feature of the input data pattern (Merkl, 1998). Each of the output nodes j is associated with a weight vector wj of the same dimension as the input data vectors and a position in the map space. The popular arrangement of output nodes is a regular spacing in a hexagonal or rectangular grid. The procedure for placing a vector from input data space onto the map is to first find the node with the spatially closest weight vector to the input data space vector. Once the closest node is located it is assigned the values from the vector taken from the data space. For Sentiment based

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 4, April 2016

classification, each vector represents an online review document, while the output node represents the sentiment category that the review is classified to. The SOM learns through self-organization of random nodes whose weights are associated to the layers of neurons. These weights are updated at every epoch during the training process. The change in weights depends upon the similarity or spatial closeness between the input data pattern and the map pattern (Michalski et al, 1999).

Step 1. Initialize weights of neurons in map.
Step 2. Randomly select a training vector xi from the corpus.
Step 3. Determine the neuron c, having the highest activity level with respect to xi.
Step 4. Update the synaptic weights of the BMU and its neighbors in order to move the BMU closer to the input vector.
Step 5. Increase time stamp t representing training iteration. If t reaches the preset maximum training time T, stop the training process; otherwise decrease α (t) ($0<\alpha$ (t).

IV. CONCLUSION

It converted a forum focused on oncology into weightedvectors to measure consumer thoughts on the drug Erlotinib using positive and negative terms alongside another list containing the side effects. This 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. Most of the user data was clustered to the area of the map linked to positive sentiment, thus reflecting the general positive view of the users. Subsequent network based modeling of the forum yielded interesting insights on the underlying information exchange among users. Modules of strongly interacting users were identified using a multiscale community detection method.

Future solutions will require more advanced detection of intersocial dynamics and its effects on the members: such interestsof study may include rankings, 'likes' of posts, and friendships. Further emphasis on context posting will require formal language dictionaries that include medical terms for specific diseases, and informal language terms ('slang') to clarify posts. Finally, different platforms will allow up-to-date information on the status of the drug in case one social platform ceases to discuss the drug.

REFERENCES

[1] A. Ng, A. Zheng, and M. Jordan, "Stable algorithms for link analysis," in Proc. SIGIR Conf. Inform. Retrieval., New Orleans, LouisianaLO, USA, 2001, pp. 258–266.
[2] 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.
[3] 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.
[4] 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.
[5] L. Dunbrack, "Pharma 2.0 – social media and pharmaceutical sales and marketing," inProc. Dec. 2005.
[6] L. Toldo, "Text mining fundamentals for business analytics," presented at the 11th Annual Text and Social Analytics Summit, Boston, MA, USA, 2013.
[7] 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: In Tech, pp. 267–297, 2010.
[8] Q. Lu. And and L. Getoor, "Link-based classification," in Proc. 20th Int. Conf. Mach. Learning, Washington, D.C., USA, 2003, pp. 496–503.
[9] W. Cornell and W. Cornell. (2013). How Data Mining Drives Pharma: Information as a Raw Material and Product.
[10]G.Angulakshmi1, Dr.R.ManickaChezian. (2014). An Analysis on Opinion Mining: Techniques and Tools.