Indian J.Sci.Res. 17(2): 183-189, 2018

ISSN: 0976-2876 (Print)
ISSN: 2250-0138(Online)

A STRUCTURE FOR ANALYZING AND DETRACTING NEGATIVE EMOTIONAL CONTAMINATION IN ONLINE SOCIAL NETWORKS

[1] Dr.Lokesh A, [2] Dr.Hema M S, [3] Dr.Nageswara Guptha

[1] Department of ISE, Sri Venkateshwara College Of Engineering, Bengaluru, Karnataka, India.,
[2] Department of Computer Science and Engineering Aurora's Scientific Technological and Research Academy, Hyderabad.
[3] Department of Computer Science and Engineering, Sri Venkateshwara College Of Engineering, Bengaluru, Karnataka.

Abstract - Now a day's social networks are being a powerful platform and also being used for spreading the negative emotion contagion. Where this type of negative emotions affecting users from different perspectives like psychologically, economically, spiritually, marketing and neuroscience. Online social networks have huge amount of data and knowledge that need to be studied using data mining techniques. In this paper, we are mainly concentrating on presenting a new framework for analyzing and detracting negative emotional contagion by making use of clustering for detecting the community where the negative emotions may spread. We are also categorizing the nodes in the network to analyze the negativity in the nodes to help decide on the best treatment. More -over, Prominent Actors in the network must be determined to help in the treatment. So finally, we are using recommender system and positive contagion the best treatment will be introduced. This framework is important to find the negative emotional contagion in different online communities and try to cure it or immunize the public against it.

Keywords - Social networks, neuroscience

I.Introduction

Ideas and information are spreading in social networks like a pathogen, each infected person effects nave friends around him. Phrases such as "social contagion" and "contagious ideas" are used to express this massive spread and infectious in different populations. Members of social networks influence not only their direct contacts but also friends' friends, up to a network distance extending beyond their closest peers. Recent studies have shown that influenza's infection rate in social networks could be as high as 25-50 % of the population and associated economic impact could reach the trillions of dollar also 13.3-17.1% of morbidity and mortality caused by depression in the United States associated by economic impact of $86 billion in the year 2000. Studies have also focused on social networks as determinants of health ranging from defining the patterns of Infectious disease spread to the propagation of behaviors such as smoking , obesity and suicidal ideation. Moreover, Emotions such as happiness and depression have been studied and found to be contagious both by media and by personal contact with strangers or friends. The popularity and massive growth of online social networks are negatively effecting users' health behavior, i.e. obesity, and loneliness "Arab Spring" and "Occupy movement" have recently showed the powerful influence of social media reported that large networking sites such as Face book or Twitter can have a significant psychological impact on our behavior, they compared people's beliefs and showed how peoples views are changing for the worse.

1.1 Existing System

In the previous approaches First tweets were obtained for a given hash tag . After that data cleaning was applied so that all the stop words like a. is ,the are removed. After the data cleaning step has taken place then the each tweet is converted into tokens. Each tweet can form a set of N tokens and then frequency computation is performed which is count of number of times a word is repeated in a given tweet statement. After that a threshold is taken from the user say T. If the frequency of any token is higher than T then the token becomes a concept. The tweets are now divided across different concepts and community is formed.

1.2 Disadvantages of Existing System

- The previous approach does not consider the sentiment analysis and also the concepts are considered based on frequency count. It does consider standard concepts.
- The previous approach does not form the classification based on negative or positive sentiments.
- The previous approach does not perform any recommendations or treatment for the users or community of users.

[1]Corresponding Author

1.3 Proposed System

In the current approach a set of HASH tags are given. For each of the HASH Tags tweets are collected. Each of tweets is analyzed by finding out the weight factor with respect to positive sentiments and negative sentiments. After performing the sentiment analysis algorithm the users are Grouped into positive community and negative community based on the classification algorithm.

For negative community alone in order to perform treatment /recommendations we divide the community into heart disease, obesity, High BP. The entire negative community users are classified into these 3 categories by applying sequence of steps like positive probability, negative probability, positive cat ratio, negative cat ratio and classified by applying a sort of positive polarity maximum, negative polarity minimum and neutral polarity maximum. The recommendations are provided for the users based on their polarity computations.

1.4 Scope of the Proposed System

The scope is limited to have the following functions

- Design and Development of Tweet Collection for Hash tags using Twitter OAuthAPI.
- Design and Development of **Positive Polarity Computation**. This is a process in which each of the tweets is separated by a delimiter. Each of the statements in the all the tweets are then compared against the positive thoughts or positive keywords listed by the data mining forums.
- Design and Development of **Group Formation** based on positive and negative polarity weight computation.
- Design and Development of Classification of **Negative Groups** in terms of Murder, Drug Supply Smuggling, Kidnap etc.

II. Literature Survey

In the paper titled "*The Simple Rules of Social Contagion*" the authors describe that It is commonly believed that information spreads between individuals like a pathogen, with each exposure by an informed friend potentially resulting in a naive individual becoming infected. However, empirical studies of social media suggest that individual response to repeated exposure to information is far more complex. As a proxy for intervention experiments, we compare user responses to multiple exposures on two different social media sites, Twitter and Digg. We show that the position of exposing messages on the user-interface strongly affects social contagion. Accounting for this visibility significantly simplifies the dynamics of social contagion. The likelihood an individual will spread information increases monotonically with exposure, while explicit feedback about how many friends have previously spread it increases the likelihood of a response. We provide a framework for unifying information visibility, divided attention, and explicit social feedback to predict the temporal dynamics of user behavior.

In the paper titled "*Origin of Peer Influence in Social Networks*" the authors describe that Social networks pervade our everyday lives: we interact, influence, and are influenced by our friends and acquaintances. With the advent of the World Wide Web, large amounts of data on social networks have become available, allowing the quantitative analysis of the distribution of information on them, including behavioral traits and fads. Recent studies of correlations among members of a social network, who exhibit the same trait, have shown that individuals influence not only their direct contacts but also friends' friends, up to a network distance extending beyond their closest peers. Here, we show how such patterns of correlations between peers emerge in networked populations. We use standard models (yet reflecting intrinsically different mechanisms) of information spreading to argue that empirically observed patterns of correlation among peers emerge naturally from a wide range of dynamics, being essentially independent of the type of information, on how it spreads, and even on the class of underlying network that interconnects individuals.

In the paper titled "*The chronicle of influenza epidemics*." the authors describe that Epidemics that were probably influenza have been reported throughout recorded history. There were 13 fairly severe epidemics during the 18th century and 12 during the 19th century. Probably 8 of these 25 were influenza pandemic. In the 20th century there have been 4 pandemic (1918/19, 1957/58, 1968/69 and 1977) due to the emergence of new subtypes of influenza A virus. The great pandemic of 1918/19 caused an estimated 20 million deaths. Between pandemics usually there have been epidemics of varying severity at intervals of one to three years and a trickle of sporadic cases every winter. The morbidity and mortality rates have varied greatly from epidemic to epidemic and from place to place during the same epidemic. Generally the morbidity has been lowest in people over 60 years of age, but, except for 1918/19, the mortality has been predominantly in old people. The epidemic behaviour of influenza has been so erratic and difficult to understand that there are still a few scientists who consider that extraterrestrial influences operate. These views are not taken seriously by most virologists but there are puzzling aspects of influenza that are not yet understood.

2.1 High Level Design

Design is one of the most important phases of software development. The design is a creative process in which a system organization is established that will satisfy the functional and non-functional system requirements. Large Systems are always are decomposed into sub-systems that provide some related set of services. The output of the design process is a description of the show how data flows through a sequence of processing steps. The data is transformed at each step before moving on to the next stage. These processing steps or transformations are program functions when Data Flow diagrams are used to document a software design. DFD diagram is composed of four elements, which are process, data flow, external entity and data store.

Figure 1: Data Flow Diagram Level 0

Fig shows the Data Flow Diagram of Level 0. As shown in the fig the input is List of Tweets. The process that is applied is Sentiment Analysis and Grouping. The outputs are Sentiment Analysis and Classification Output.

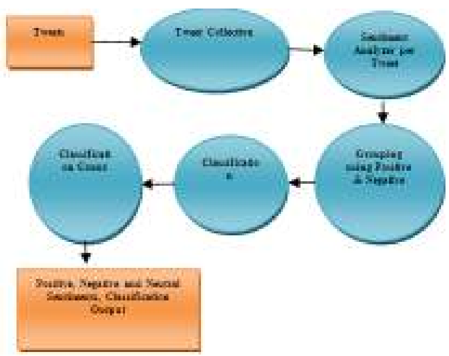


Figure 2: Data Flow Diagram Level 1

Fig shows the Data Flow Diagram of Level 1. As shown in the fig the input is List of Tweets. The process that is applied is Tweet Collection, Sentiment Analysis per Tweet, Grouping using Positive & Negative, Classification and Classification Count. The output is the positive, negative and neutral.

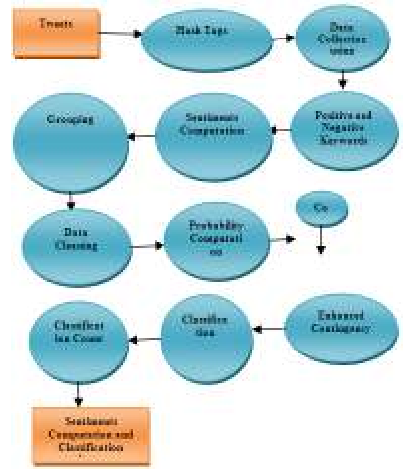
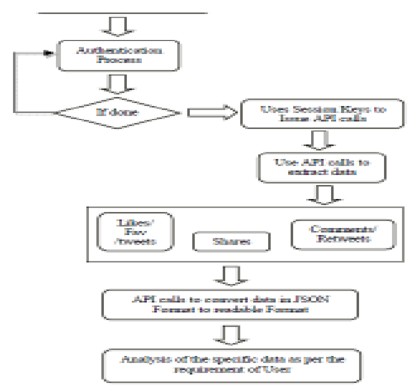


Figure 3: Data Flow Diagram Level 3

Fig shows the Data Flow Diagram of Level 3. As shown in the fig the input is List of Tweets. The process that is applied is Tweet Collection, Sentiment Analysis, Data Cleaning, Probability Computation, Contingency Computation, Enhance Contingency, Classification and Classification Count. The output is the Sentiments Computation and Classification Count.

2.2 System Design



User Authentication Flowchart

Authenticated requests are must to access the API's of SNS's. Each request must be signed with valid user credentials *OAuth.*It is an Open authorization protocol specification defined by IETF OAuth WG (Working Group) which enables applications to access each other's data . It is an open standard for authentication, adopted by Twitter to provide access to protected information and the process is carried out using a three-way handshake.

The client 'gets a token' from part of Web Server i.e. Auth Server and then 'uses the token' to authenticate to another part of Web Server i.e. Resource Provider, which is the data the client desires to obtain or manipulate. OAuth provides a method of third party authentication that allows Web services to share data through their APIs.

Data Collection Using Twitter

The data collection using twitter provides a set of hash tags. After entering the hash tags the hash tags are stored and then data is collected for each of colleges using the Hash Tag. The authentication process makes use of OAuth Authentication API in which secret key and token is provided. After that the request is send to twitter and if the authentication is successful the tweets are collected for the set of hash tags across various colleges.

The tweet storage based on Tweets is a process of storing the data about the tweets into the relational storage in terms of *(TwitterId, TwitterDesc, UserId)*. Twitter Id is unique Id associated with the tweet, *TwitterDesc* is the actual tweet and *UserId* is the Id associated with the user.

Data Cleaning Algorithm

The Data Cleaning algorithm is responsible for removal of stop words. Each of tweet is cleaned by removing the stopwords from tweet.

These are the set of words which do not have any specific meaning. The data mining forum has defined set of keywords. Stop words are words which are filtered out before or after processing of natural language data (text). There is not one definite list of stop words which all tools use and such a filter is not always used. The list of stop words used in the algorithm are as follows

a,able,about,across,after,all,almost,also,am,among,an,and,any,are,as,at,be,because,been,but,by,can,cannot,could,dear,did,do,does,either,else,ever,every,for,from,get,got,had,has,have,he,her,hers,him,his,how,however,i,if,in,into,is,it,its,just,least,let,like,likely,may,me,might,most,must,my,neither,no,nor,not,of,off,often,on,only,or,other,our,own,rather,said,say,says,she,should,since,so,some,than,that,the,their,them,then,there,these,they,this,tis,to,too,twas,us,wants,was,we,were,what,when,where,which,while,who,whom,why,will,with,would,yet,you,your

Data Cleaning is used for removing the stop words from each of the tweets and clean them. After the data cleaning process is completed the clean data can be represented as a set *(CleanId , CleanData ,UserId)*. *CleanId* is the unique Id associated with the Tweet, *CleanData* is the clean data after removal of clean data and *UserId* is the unique Id associated with the user.

III. Implementation

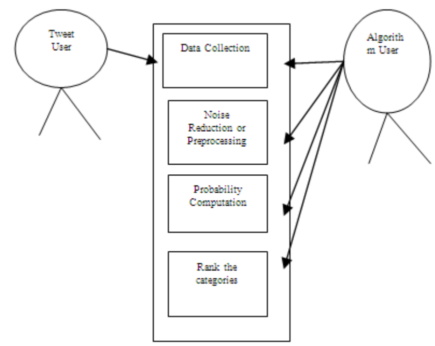


Fig shows the Use Case Diagram as shown in the figure there are two kinds of users namely Tweet User and Algorithm User and the functions performed by them.

3.1 Activity Diagram

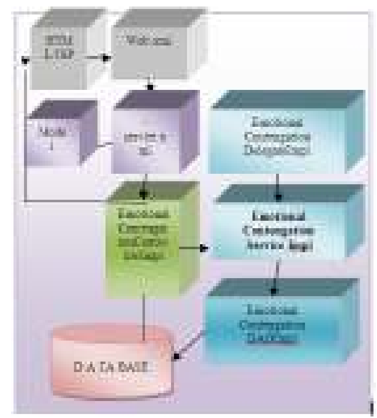


Figure 4: Tweet Activity Diagram

The above figure gives description about the system architecture which is followed in the industries in order to a development of any routing software.

The figure shows that the user interface is designed in the HTML/JSP pages and then the request goes to the web container and web container verifies the request in the web.xml file by looking first into the url pattern and then it goes to the Controller name and then it searches for the corresponding Controller name in the Controller tag and looks into the Controller class and creates an object of TweetController and then the TweetController will delegate its job to Request Processor.

The request processor will look for the action to which must be called in looked up in the -servlet.xml and corresponding form is called and then the action is called. The action class will then call the delegate , then the delegate calls the service and service calls the Data Access layer and results goes exactly in the opposite way and the resultant JSP page is loaded

Model This is the Plain Old Java Object which will have the getters and setters and setters gets automatically called and data the user has entered will be available.

Controller This is the class which is used to fetch the user entered data and then processes it and calls the delegate layer and obtains the results.

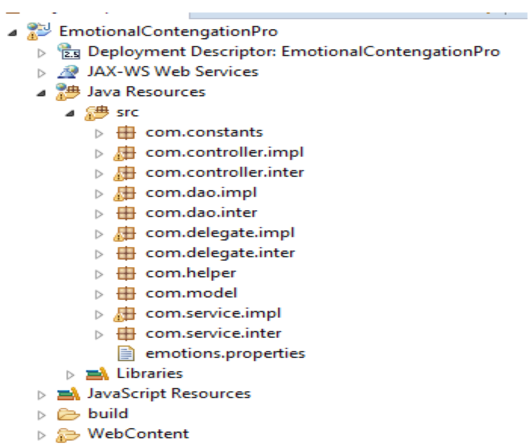
Delegate Delegate is the layer which contains nothing but call to an appropriate service.

Service This is the layer which is responsible for entire algorithmic implementation. This is the layer which contains the heavy weight implementation of entire algorithms. Future the service would require the help of Data Access Layer for some operations and many other helper classes.

Data Access Layer This is the layer which deals with only the CRUD operations namely Create, Retrieve, Update and Delete. It has no other usage. This layer has been used in order to fetch the data from the routing tables.

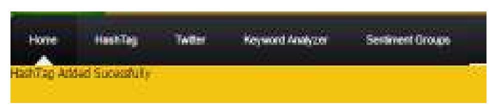
Database This is the place where all the tables would have been placed have been placed.

III.2 Packaging Diagram

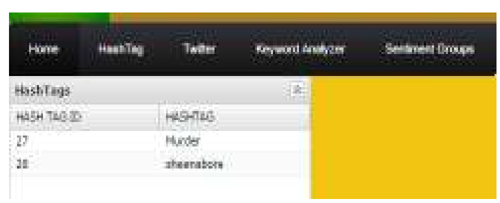


3.3. Class Diagram

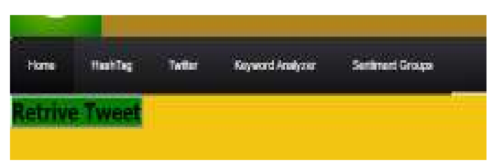
IV. Results



View Hashtags



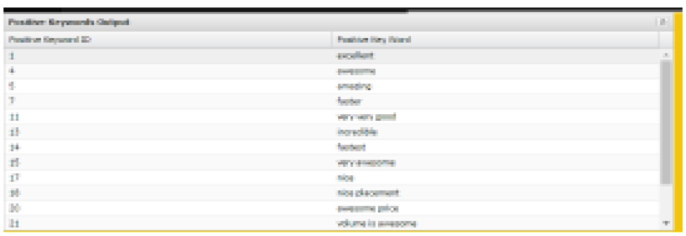
Twitter Data Collection



View Tweets

Positive Keyword Addition

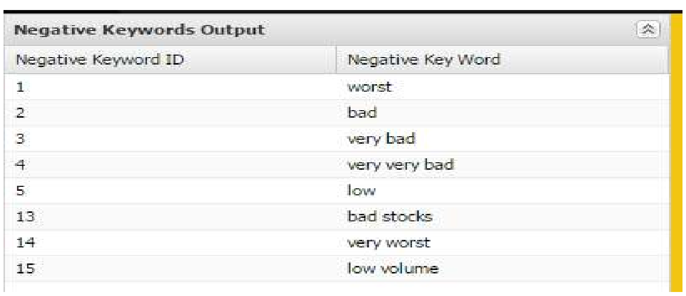
View Positive



Positive Keywords Output

Positive Keyword ID	Positive Key Word
1	excellent
4	awesome
5	amazing
7	faster
11	very very good
13	incredible
14	fastest
15	very awesome
17	nice
18	nice placement
20	awesome price
21	volume is awesome

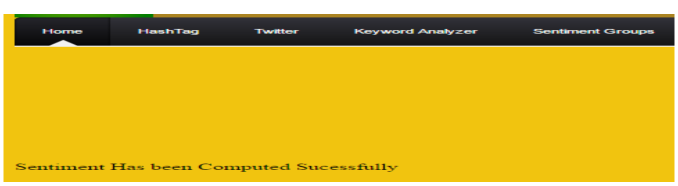
View Negative



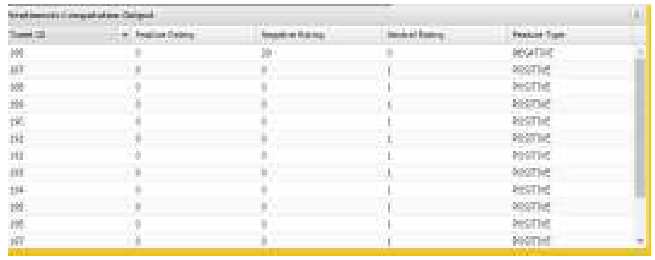
Negative Keywords Output

Negative Keyword ID	Negative Key Word
1	worst
2	bad
3	very bad
4	very very bad
5	low
13	bad stocks
14	very worst
15	low volume

Perform Sentiments



View Sentiments



V.Conclusion & Future Scope

Framework aids the E-crime department to identify suspicious words from cyber messages and trace the suspected culprits. Currently existing Instant Messengers and Social Networking Sites lack these features of capturing significant suspicious patterns of threat activity from dynamic messages and find relationships among people, places and things during online chat, as criminals have adapted to it.

In this paper the tweet are retrieved from the social networking application namely Twitter, after finding the list of tweets Positive, Negative and Neutral Sentiments. The tweets are grouped into 3 types namely Positive and Negative Groups. For the Specific Group of tweets namely Contingency, Enhance Contingency, Classification and finally Classification Count

1) The paper can be future improved by adding more tweets for each of the hash tags.

2) The framework can be used to obtain web page.

References

[1] N. O. Hodas and K. Lerman, The Simple Rules of Social Contagion, *Sci. Rep.*, vol. 4, Mar. 3014.

[3] F. L. Pinheiro, M. D. Santos, F. C. Santos, and J. M. Pacheco, Origin of Peer Influence in Social Networks, *Phys. Rev. Lett.*, vol.113, no. 9, p. 098703, Mar. 3014.

[3] C. W. Potter, Chronicle of influenza pandemics, *Textb. Influenza Oxf. Blackwell Sci.*, vol. 3, 1998

[4] M. Ventresca and D. Aleman, Evaluation of strategies to mitigate contagion spread using social network characteristics, *Soc. Netw.*,vol. 35, no. 1, pp. 7588, 3013.

[5] C. N. Dulmus and J. S. Wodarski, Major depressive Disorder and dysthymic disorder, *Handb. Empir. Soc. Work Pract. Vol. One N.Y. Wiley*, pp. 373385, 1998.

[6] J. N. Rosenquist, J. H. Fowler, and N. A. Christakis, Social network determinants of depression, *Mol. Psychiatry*, vol. 16, no.3, pp. 373381, Mar. 3011.

[7] K. P. Smith and N. A. Christakis, Social Networks and Health, *Annu. Rev. Sociol*., vol. 34, no. 1, pp. 405439, Aug. 3008

[8] M. J. Keeling and K. T. Eames, Networks and epidemic models, *J. R. Soc. Interface*, vol. 3, no. 4, pp. 395307, 3005

[9] N. A. Christakis and J. H. Fowler, The collective dynamics of smoking in a large social network, *N. Engl. J. Med.*, vol. 358, no.31, pp. 3349-3358, 3008

[10] N. A. Christakis and J. H. Fowler, The spread of obesity in a large social network over 33 years, *N. Engl. J. Med.*, vol. 357,no. 4, pp. 370379, 3007

[11] A. L. Hill, D. G. Rand, M. A. Nowak, and N. A. Christakis, Emotions as infectious diseases in a large social network: the SISa model, *Proc. R. Soc. B Biol. Sci.*, vol. 377, no. 1701, pp. 38373835, Dec. 3010.

[13] P.S.Bearman and J.Moody, Suicide and friendships among American adolescents,*Am. J.Public Health*, vol.94, no.1, pp.8995, 3004

[13] T. E. Joiner and J. Katz, Contagion of Depressive Symptoms and Mood: Metaanalytic Review and Explanations From Cognitive,Behavioral, and Interpersonal Viewpoints, *Clin. Psychol. Sci. Pract.,* vol. 6, no. 3, pp. 149164, 1999

[14] J. H. Fowler and N. A. Christakis, Dynamic spread of happiness in a large social network: longitudinal analysis over 30 years in the Framingham Heart Study, *Bmj*, vol. 337, 3008.