ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Classification Based Diffusion Analysis in Social Network

Manali Wadekar, Prashant Itankar, Ankit Sanghavi

M.E Student, Department of Computer Engineering, ARMIET Shahapur, Mumbai University, Mumbai, Maharashtra, India

Associate Professor, Department of Computer Engineering, DMCE, Airoli, Mumbai University, Mumbai, Maharashtra, India

Head of the Department, Department of Computer Engineering, ARMIET Shahapur, Maharashtra, India

ABSTRACT: Sentiment Analysis has been a hot research topic in current era. Emotion classification is more detailed sentiment analysis which cares about more than the divergence of sentiment. In this paper addresses the problem of accurately classifying the sentiment in posts from micro-blogs such as Twitter. As Twitter gains fame, it becomes more useful to study trends and sentiment of its users towards various phenomena. I learn the diffusion strength in different triadic structures by using sentiment analysis. The learned diffusion strength is evaluated through sentiment classification to improve the applications of follower maximization and followee recommendation, which are specific instances of influence maximization. Our tentative results reveal that incorporating diffusion patterns can indeed lead to statistically major improvements over the performance of several substitute methods, which demonstrates the effect of the discovered patterns and diffusion model.
The project addresses this problem by gathering data, pre-processing perform on the data, and analyzing the required data using different techniques of machine learning to classify them by sentiment as either negative, positive, or neutral. I investigate different methods for pre-processing the micro-blogging posts and use various Naive Bayes classification and feature selection methods to determine the best approach. In addition, address the problem of building classifier for different classification.

KEYWORDS: Diffusion, Social Network.

I. INTRODUCTION

In the era of 21st century technology plays vital role. People are more prone to use social networking sites. Social network is most emerging platform for marketing and advertisement. An individual's choice to adopt a new behaviour often depends on the distribution of similar choices the individual observes among her society, be they friends, colleagues, or acquaintances. This may be determined by underlying network externalities, as in a decision to use a new product such as a new home appliance or new accessories, where the benefits of the new products are larger when more of an agent's acquaintance has adopted the products. It may also be an artefact of simple learning processes, where the probability that an individual learns about a new behaviour or its benefits is increasing in the number of neighbours who have adopted the behaviour. For example, decisions concerning whether to go to a particular movie or restaurant, or to buy a new product, provide examples of situations in which information acquired through friends and their behaviour are important. Of course, there are many other likely channels by which peer decisions may have major impact on individual behaviour. The starting point of our analysis is the observation that in all such environments, the extent to which a new behaviour influence society depends not only on its virtual attractiveness or value of money, but also on the underlying social structure.

In a social Network such as Twitter, users "following" behaviours form the "following" link, which is elementary to the formation of a network structure. Following link observed to be correlated. For Example, when a user Atharv follows user Sarthak, this creates a chance for Atharv's follower Devanshi to discover Sarthak, where Atharv, Devanshi

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

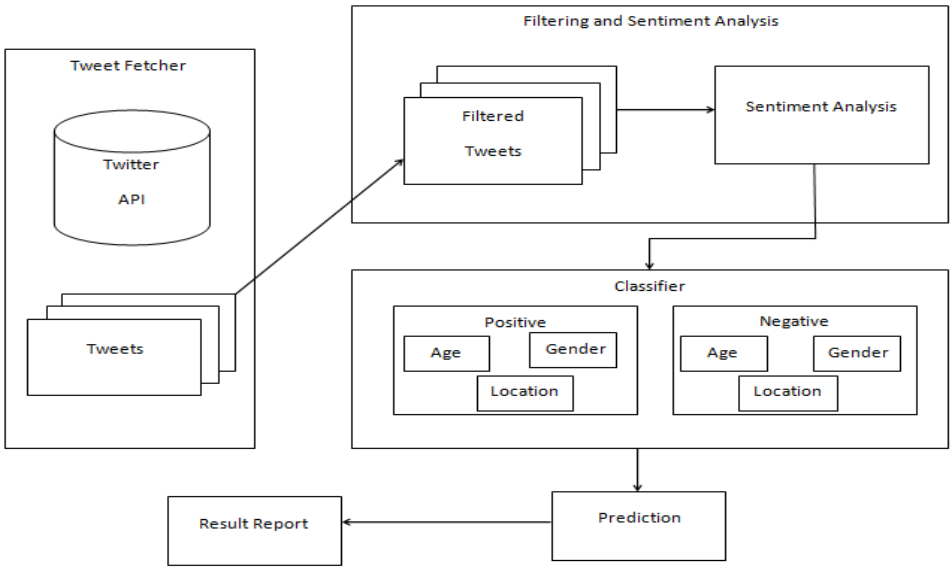
and Sarthak form a basic triadic structure. In this paper we will study the diffusion in micro blogging network with different classification parameter. I will analyze our results for study of real time applications.

The main purpose of this diffusion analysis to know followers and followee classification by studying phenomenon of diffusion on Twitter network. These algorithms are not just to study diffusion but also to make classification of different followers in the network structure. This algorithm based on two parameter such as sentiment analysis as well as classification phenomena which provide classification based on age, gender and location.[1]

II. LITERATURE SURVEY

The author in [2] discuss the diffusion analysis of events with respect to space and time using retweet interaction networks between users. Therefore, space and time analysis for distribution of event related tweets on Twitter gives us an insight into the characteristics exhibited by different event discussion phenomenon in social media networks. Erick Stattner, Reynald Eugenie, Martine Collard, May 13-15 2015 Athens [3] in this paper we Understand how the information spread among the network structure. Mostly, categorize in which conditions a person retweets for particular event. One of the important factor is to understand how information diffuse through social networks and what can positively or negatively affect it. In paper [4], author Masud Karim, Rashedur M. Rahman explain how the naives classifier classify the gathered data so we implement naive bayes classifier to classify the followers to identify the various category. Data gathering for this task involved more effort than expected and required hand-tagging posts for sentiment in relative to a query. We use Twitter API to collect data for various queries ranging from movies and actors to sports teams and politicians. By mainly gathering posts related to current topics. [5] "Sentiment Analysis: Capturing Favorability Using Natural Language Processing" propose the sentiment analysis algorithm use natural language processing to classify documents as positive , neutral or negative. Developers and scientist working on big data write software which feeds documents into the algorithm and stores the results in a way which is useful for clients to use and understand. [6] Walaa Medhat, Ahmed Hassan , Hoda Korashy , "Sentiment analysis algorithms and applications: A survey", in Ain Shams Engineering Journal (2014) 5, 1093–1113 provides sentiments classification techniques on product and process which help us to study text mining.

III. SYSTEM ARCHITECTURE



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

IV. PROPOSED ALGORITHM

A. *DESIGN CONSIDERATIONS:*

- Initial set of followers for a any topic are taken from real time data.
- Collection of number of followers only of a particular scenario for which we are testing.
- Keep track of followers for dynamic database since dynamic data updated at an every instant.

B. *DESCRIPTION OF THE PROPOSED ALGORITHM:*

Aim of the proposed algorithm is to identify people as positive or negative tweets from their appearance. The proposed algorithm is consists of two main steps.

Steps for Sentiment Analysis:

Step 1: Text Collection: In this steps we collect different text for generating data which is first step of data gathering.This gathered data can be static or it can be updated at every instant called as dynamic data.

Step 2: Pre-processing: This step involves preparing gathered text in more structured form for representation

Step 2.1) Tokenization: Tokenization is used to identify all words in a given text.

Step 2.2) Data Filtering: People use a lot of casual language on twitter. For example, 'sad' is used in the form of 'saaad' or can be present using smiley. Though this implies the same word ‘sad’, the system considers these as two different terms. To pick up this and make words more analogous to standard words, such sets of recurring letters are replaced by two occurrences. Thus ‘saaad’ would be replaced by ‘sad’.

Step 2.3) Stop Word Removal: This step involved to eliminate the words that occur frequently such as article, prepositions, conjunction and adverbs. These words depend on lingo of the text in questions. For example, words like the, and, before, while, and so on do not contribute to the sentiment.

Step 2. 4) Stemming: In method of information retrieval, stemming is the process of tumbling a word to its root form. For example, talking, talkative, talked all these words are derived from the root word talk. Hence, the stemmed form of all the above words is talk.

Steps for classification:

Generate decision tree from the training tuples of data partition D

Input:

1. Data partition D which is set of training tuples & their associated class labels;
2. Attribute List and set of Candidate attributes.
3. Attribute Selection Method a procedure to determine splitting criteria that “best” partition data tuples into individual cases. The criteria consist of splitting attribute & possibly split point or splitting subset.

Output:

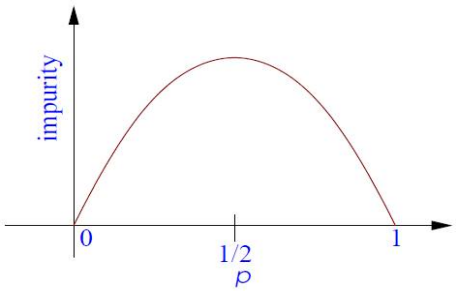
Decision Tree

V. PSEUDO CODE

Step 1: choose rule to split on

Choosing best rule to Split On: choose rule that leads to greatest increase in “purity”

(p = fraction of positive examples)



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

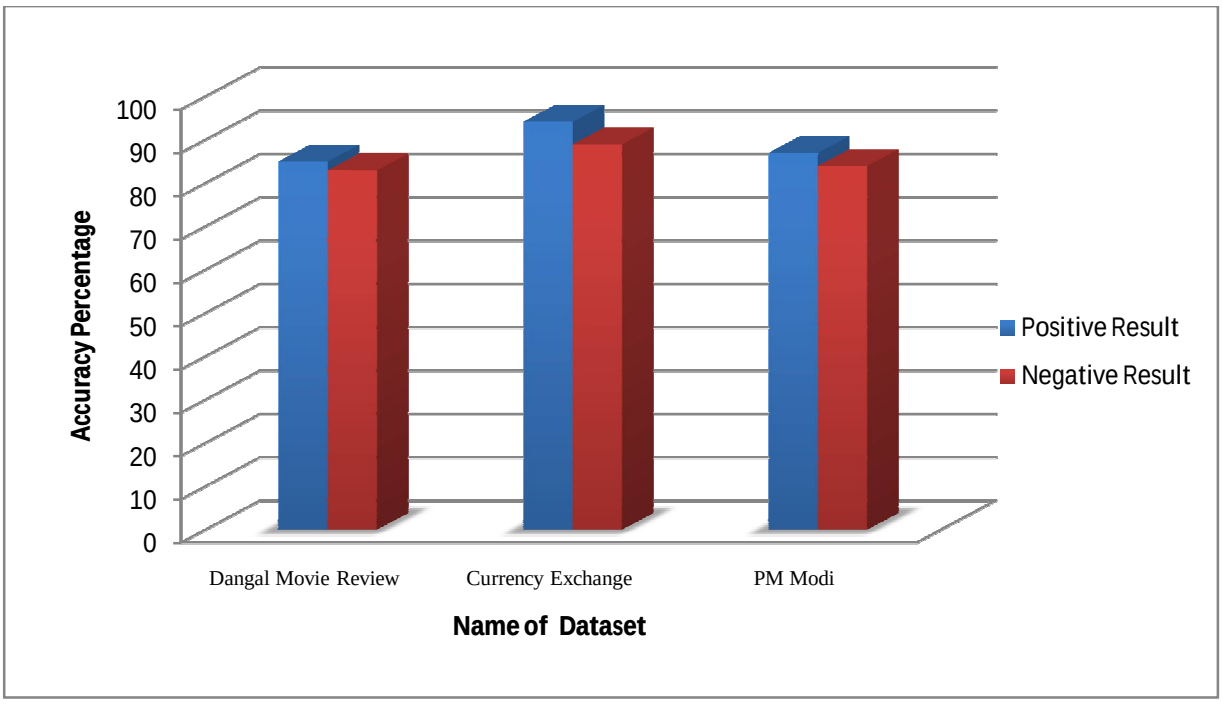
Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

commonly used impurity measures:
entropy:: $p \ln p - (1 - p) \ln(1 - p)$
Gini index:: $p(1 - p)$
Step 2 : divide data using splitting rule into disjoint subsets, as shown below
Step 3: repeat recursively for each subset
Step 4 :stop when leaves are (almost) "pure"
Step 5: End.

VI. SIMULATION RESULTS

The simulation studies involve the study for diffusion of "following" link with small dataset taken from twitter data. The proposed algorithm is implemented with JAVA.The Proposed algorithm is tested with tweets on #CurrencyExchange, #DangallMovieReview and #PMModi. For the currency exchanged event out of 5000 tweets 4200 were found positive and 800 were found negative with sentiment analysis. The trained model was successful in predicting 3956 positive tweets correctly giving accuracy of 94.19 percent. I have tested results same as above for PM Modi and Dangal movie review giving sentiment analysis accuracy is varying 86 to 94 percent.



VII. CONCLUSION AND FUTURE WORK

The simulation results showed that the proposed algorithm performs better with the total transmission energy metric than the maximum number of hops metric. The proposed algorithm provides energy efficient path for data transmission and maximizes the lifetime of entire network. As the performance of the proposed algorithm is analyzed between two metrics in future with some modifications in design considerations the performance of the proposed algorithm can be compared with other energy efficient algorithm. We have used very small network of 5 nodes, as number of nodes increases the complexity will increase. We can increase the number of nodes and analyze the performance.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

REFERENCES

1. Manali Wadekar, Prashant Itankar and Ankit Sanghavi, "Classification based diffusion analysis in social network", International Journal of Emerging Trends & Technology in computer Science.
2. Purva Pruthi, Anu Yadav, Farheen Abbasi and Durga Toshniwal, "How has Twitter changed the Event Discussion Scenario?: A Spatio-Temporal Diffusion Analysis", Indian Institute of Technology Roorkee, Uttarakhand, India.
3. Erick Stattner, Reynald Eugenie and Martine Collard , "How do we Spread on Twitter?", LAMIA Laboratory University of the French West Indies and Guiana, France {erick.stattner, reynald.eugenie, martine.collard}@univ-ag.fr May 13-15 2015 Athens, Greece.
4. Masud Karim, Rashedur M. Rahman, "Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing", Journal of Software Engineering and Applications, 2013, 6, 196-206.
5. Tetsuya Nasukawa , Jeonghee Yi, "Sentiment Analysis: Capturing Favorability Using Natural Language Processing", IBM Research, Tokyo Research Laboratory.
6. Walaa Medhat, Ahmed Hassan , Hoda Korashy , "Sentiment analysis algorithms and applications: A survey", In Ain Shams Engineering Journal (2014) 5, 1093–1113.

BIOGRAPHY

Ms. Manali Mahadev Wadekar has received B.E. in Computer Engineering from Mumbai University, Mumbai, India in 2011. Currently she is pursuing M.E. in Computer Engineering from Mumbai University, Mumbai, India.