ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

Prediction of User Opinion in Microblog

Tabrez Khan[1], Javed Khan Sheikh[2], Rehal Qureshi[3],

Nouman Khatib[4], Priya Goankar[5], Fauziya Khan[6], Siddiqua Shaikh[7]

Assistant Professor, Computer Engineering, AIKTC, Navi Mumbai, India[1,2,3]

BE Student, Computer Engineering, AIKTC, Navi Mumbai, India[4,5,6,7]

Abstract: In this paper, user's opinion on social media in the form of comments will be analysed to draw conclusion on preference topics. This will be done by using techniques of collaborative filtering, natural language processing and data mining methodology. This will help social media users to determine the conclusion about comments. Analysis will enable user for effective and fast decision making on specific topic. Also Users can find reviews of any product, movie, social issue, political review etc. at single blog itself.

Keywords: Natural language processing,Social Media ,filtering, data mining.

I. INTRODUCTION

Microblogging has become very popular and changed the way people interact with each other. Social media has exploded as a category of online discourse where people create content and share it at a prodigious rate. However, there is an overload problem of information. Besides regular posts, commercial posts also exist for promotion and branding in microblogging websites. The rapid updated streams of microblogging posts provide an interesting opportunity for harnessing data into a form that allows for specific predictions about particular outcomes.

Our objectives for this paper are as follows First, we will posts those topics that will create attention then user will give his/her opinion on their preferred topic in comment box. and these opinion will be stored in database.

Filtering will be applied on comments which will remove those words that are not required or useless. And Those comments will be analysed using sentiments and common words lists to determine user opinion regarding the topic.

Such that it shows how positive and negative opinions propagate .For a bad comment, the initial reviews might be enough for users and specially for corporate user, to know the performance, while on the other hand positive reviews will encourage the users.

Further representation of this paper are as follows: Next we give a brief description about how we got the idea about this paper in section II. In section III we have describes the working that how we are able to predict user opinions. In section IV we describe our system architecture and its description is explained in section V and further market potential and future work are described.

II. LITERATURE REVIEW

There are many studies related on this topic In this section we describe from where we have gathered ideas for perparing our paper.

Nutshell about the related work are as follows:

[1] "Exploiting Social and Topical Contex for Predicting User Preferences in Microblogging " by YE WU every day hundreds of people post their statuses with 140-character limit messages User preference is defined from two aspects: user interest and user opinion.

It focus on solving this problem with a Social context and Topical context incorporated matrix factorization method.Demonstrate that both social context and topical context are effective in improving the user-topic opinion prediction performance.

[2] "A user opinion and metadata mining scheme for predicting box office performance of movies in the social network environment" by DAEHOON KIM, DAEYONGKIM , EENJUN HWANG and HONG-GU CHOI This paper describes how to predict the box office performance on the basis of the public opinion and other marketing properties.

Its shows user comments and marketing properties together leads to better prediction accuracy.The drawbacks of this scheme are, it can handle English comments only for a movies. It can predict box office hit according to the four categories i.e., stars, director, writer, and their past works but for more practical utilization and better accuracy some more categories sholuld be considered.

[3] "Predicting User-Topic Opinions in Twitter with Social and Topical Context." by FUJI REN and YE WU describes experimental results on a real-world Twitter data set. This framework depicts the state-of-the-art using collaborative filtering methods, and demonstrate that both social and topical context are effective in improving the user-topic opinion prediction performance.

The data generated in Twitter is thus regarded as a resource providing individuals' spontaneous emotional information about their own opinion.Some of the weakness encountered are detecting the opinions of the user more accrately,hashtags,comments in multi-language are not supported.

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

III. WORKING

In this section, we will introduce basic technical flow, Over here we explain that how can we analyse the comments by using sentiments and common words lists and these comments are extracted from social media.

1. Initially user will give his/her opinion on their preferred topic in comment box.
2. These comments will be stored in database.
3. Filtering will be applied on comments which will filtered all those words which are required for analysis.
4. Those comments will be analysed using sentiments and common words lists to determine user opinion regarding the topic.

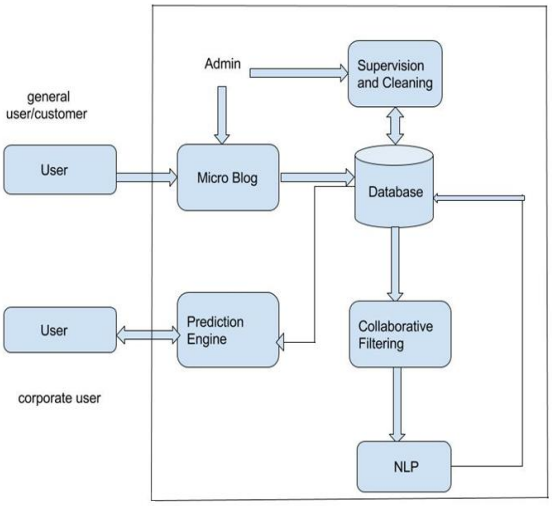
Example

A company can use this analysed opinions of public for feedback about its products in comments from microblog . Such that, opinion mining helps to collect information about the positive and negative aspects. then, the positive and negative opinions obtained about a particular product are recommended to the users.

IV. SYSTEM ARCHITECTURE

In this section we give a brief description of the technical flow related to the use of technology that are Natural Language Processing ,data filtering by collaborative method , data mining and prediction engine.

The figure shown next represents a technical flow under which the comments will undergo



In the next section we describe a brief architecture related to the system architecture flow as described in above figure

V. ARCHITECTURE DESCRIPTION

System Architecture is divided into three authorities. For implementing This project we are using combination natural language processing and data mining technologies.

Natural language processing **(NLP)** is a component of artificial intelligence, and computational linguistics concerned with the interactions between computers and human (natural) languages. As such, NLP is related to the area of human computer interaction.

Data mining sometimes called data or knowledge discovery is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used. It is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions.

Filtering data means to set conditions in such a way that only certain data is displayed. It is done to make things easier to focus on specific information in a large database or table of data

System Architecture is divided into three authorities.

Client side-represents the part, where users give their opinion on specific topic and comment will be sent to the server's database.

Server side-includes processing of user opinion by filtering user's opinions. This opinion is an important aspect in this project. Comments can be positive or negative. These comments will undergo filtrations in which unnecessary words will be filtered or removed. Filtered words are in unstructured form. These words should be stored in structured format which can done using natural language processing.

Admin Side- Admin takes responsibility for site reactivity it also takes responsibility pro activity.

There are some specific tasks performed by administrator such as:

1. Supervising the contents of database.
2. Updating illegal comments.
3. Filtering out illegal comments.

VI. MARKET POTENTIAL

Company has a knack for getting bloggers in order to have maximum idea for launching their products.This can be done by recording ones daily life, opinions and demonstrating ideas via text which is done in microbloging.

Fig 2: Graphical representation of user opinion on review

DOI 10.17148/IJARCCE.2015.410109

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 10, October 2015

Figure Reference: http://searchengineland.com/88-consumers-trust-online-reviews-much-personal-recommendations-195803

Market Potential is the entire size of the market for a product at a specific time.

This graph represents the percentage of customer's who valued online review over the years. As shown in this graph from 2011-2014 the demand of review have been increasing as the co-operate customer requirement is also increasing for their product manufacturing in order to satisfy the requirement of the users.

There are existing software in market like "Exploiting Social and Topical Context for Predicting User Preference in Micro blogging"," A user opinion and metadata mining scheme for predicting box office performance of movies in the social network environment" .it combines user comments made on its trailer and marketing properties including information on the leading stars, director, writer, and their past works for better prediction accuracy. The box office record prediction system was trained using information on released movies as the ground truth, but there are some draw backs which will be overcome by our project.

In this existing project sentiments of each comment are determine or identified using the sentiment word and common use word lists and then they calculate the percentages of positive comments and negative comments over total comments, respectively. Their difference becomes the sentiment score of the comments, which ranges from −1 to 1.The problem in analyzing this type of comments is shown below with an for example:

For example: Acting was good, but the movie was bore

Attributes: Good, bore

Weight-age: good (positive word) =1, bore(negative word)= -1

Result: 1-1=0

Predicting such diplomatic statements is bit difficult.

Our project stands different to other software due to following reasons.

- Opinion Accuracy
- Users can find reviews of any product, movie, social issue, political review etc. at single blog itself.
- It is efficient.

VII. CONCLUSION AND FUTURE SCOPE

In this project, we focus on a challenging problem of predicting users' opinions toward a particular topic which we define as user-topic opinion prediction. The main contributions of this project are as follows:

(1) We find out who has what opinion of a specific topic through micro blogging and these opinions will be sent to the server's database. We believe that predicting individual's feeling about a given target is important to be used to various applications.

(2)All the contents of blog are handled by Admin such as updating of posts, supervising the contents of database, updating illegal comments and Filtering out illegal comments.

(3) The review of the comments includes processing of the user opinions by filtering the key words, finding the user emotions by Natural language processing and user will get the final review about the project

We satisfy the requirement of customer and brands, by providing an improved performance of user- topic opinion prediction.Some limitations we encountered in this paper which is giving us a directions for future scope.

This project currently can handle English language only. We need to extend it to support multilingual comments. In that way, we can gather comments from non- English speaking users together and extract more diverse opinions for prediction.We suppose that there should be explicit states to describe users' opinions towards large number of topics than positive and negative. To predict multiple opinion/ emotion states users have of topics will be interesting tolearnAn effective and efficient approach to detect topics from the user-generated data is in demand and also be an interesting to learn in the future.

ACKNOWLEDGMENT

First of all we are thankful to Anjuman -I- Islam Kalsekar Technical Campus.We are also grateful to H.O.D **Prof. Tabrez Khan** ,guide **Prof. Javed Sheikh** and co-guide **Prof.Rehal Qureshi** who devoted their time and knowledge for the implementation of this project. Without their superior knowledge and experience, the Project would like in quality of outcomes, and thus their support has been essential.

REFERENCES

[1] Fuji Ren, "Predicting User-Topic Opinions in Twitter with Social and Topical Context", Vol.4 No. 4, October-December 2013
[2] Daehoon Kim, Daeyong Kim, Eenjun Hwangand and Hong-Gu-Choi, "A user opinion and metadata mining scheme for predicting box office performance of movies in the social network environment", Vol. 19, 14 Nov 2013.
[3] Sheng Yu and Subhash Kak, "A survey of Perdiction Using Social Media", http://www.arxiv.org/pdf/1203.1647.pdf [Accessed on Oct 10,2015]
[4] B. Pang and L. Lee, "Sentiment Analysis and Opinion Mining," Foundations and Trends in Information Retrieval, vol. 1, pp. 1-135, 2008.
[5] A. Go, R. Bhayani, and L. Huang, "Twitter Sentiment Classification Using Distant Supervision," technical report Stanford Univ.,2009.