Research Paper | Computer Science | E-ISSN : 2454-9916 | Volume : 2 | Issue : 1 | Jan 2016

USER SENTIMENTS ANALYSIS IN TWITTER WITH SOCIAL CONTEXT

* Swapnil S. Nanaware

UG Student, Computer Department, Trinity Academy College of Engineering, Pune, India - 411048.

*Corresponding Author

ABSTRACT

Users can share their emotions in more suitable format with the help of micro blogging services like twitter. Twitter provides information about individual's real-time feelings through the data resources provided by individuals. The important task is to extract user's tweets and perform an examination and survey. However, this extracted information will helpful to make prediction about user's opinion towards specific topics. As there are tremendous amount of tweets available on micro blogging services. It is very difficult to user, so the major challenge is to analyze all tweets in short time. In this paper we mainly focus on solving this problem with the naive Bayes technique. This paper is attempted to obtain polarity of individual's opinion used for opinion and sentiment analysis.

KEYWORDS: Twitter, Opinion Mining, Sentiments, Polarity, Naïve Bayes Classifiers, Feature Extraction Technique.

Introduction:

Electronic learning has become faster and very muchconvenient due to worldwide and availability of the internet. Thecustomer's reviews are increasing various in number on variousproducts. These large numbers of reviews are beneficial tomanufacturers and organization to improve their quality as well as response. It is a complex problem for users to read all reviews to make a better decision ofpurchasing. It is helpful available customer reviewsfor popular products from various product review sites ofcustomer. There has been a lot of research activity in the areas ofopinion mining and sentiment analysis. Researchers are verymuch interested in performing opinion mining and sentimentanalysis because of the increased availability of machinelearning techniques used in natural language processing,retrieval of information and the growth of online customer's

Review-aggregation.

Problem definition:

With popular micro-blogging services like Twitter, users are able to online share their real-timefeelings in a more convenient way. The user generated data in Twitter is thus regarded as a resource providing individual's spontaneous emotional information, and has attracted much attention of researchers.Prior work has measured the sentiment expressions in user's tweets and then performed various analysisand learning. In this paper, we mainly focus on solving thisproblem with a Social context and Topical context incorporated Matrix Factorization framework. Theexperimental results on a real-world Twitter data set show that this framework outperforms the state- of-the-art collaborative filtering methods, and demonstrate that both social context and topical context are effective in improving the user-topic opinion prediction performance.

Literature survey:

In last decades there are various classification techniques ofreview available for deciding polarity of reviews from thereviews. It was firstly presented in [1], where polarity ofreviews has been used to make an improved decision ofpurchasing.

Existing system:

1. The performance of opinion mining in determining the orientations or polarity is evaluated by calculating various metrics like precision, recall and F-measure.By Rudy the overview of the work done in the task of opinion mining and its orientations is discussed where for movie review data mining technique is used as SVM which provides efficiency of 89% and by Gang li using k-means clustering performance is 78%.

2. The data source is concerned, a huge amount of work has been done on movie and product reviews to determine the opinion orientations. The Internet Movie Database is used for movie reviews and product reviews are taken from Amazon.com. Movie review is a more challenging application than many other types of review mining.

3. By Gam gam for amazon review data mining technique is maximum entropy is used which gives precision 72%,recall 78% f-measure is 75%.

4. The challenges of movie review based on the factual information which is always mixed with real-life data and mocking words are used in writing movie reviews. The Product review domain considerably differs from movie review domain because of the following reasons.

5. One reason is that there are feature specific comments in product reviews because people may like some features and dislike others. Thus reviews consists opinions orientations in the text, which is a difficult one to classifying opinion orientation of review as positive or negative. Following feature specific reviews occur less often in movie reviews.

6. Second reason is that there are a lot of comparative sentences in product reviews and people discuss about other products in reviews

Proposed system:

1. Comparing with the content of other sophisticated social media, the improvisatory short messages on micro-blogging are easier to obtain and more likely to reflect individuals spontaneous emotions. Twitter,1 as one of the most famous micro-blogging, hundreds of millions of people freely express how

they are feeling about breaking news, public figures, hot products, or just daily things on it in 140-character limit tweets (the messages posted by Twitter users) every day. General, the subjective feelings about particular matters could be defined as individual's opinions, which are considered to be the result of emotion and play an important role during the decision-making process most of the time.

2. With popular micro-blogging services like Twitter, users are able to online share their real-time feelings in a more convenient way.

Materials and Methods:
External Interface Requirements:

USER INTERFACE SYSTEMS CAN BE BROADLY CLASSIFIED AS:

1. User Initiated Interface the user is in charge, controlling the progress of the user/computer dialogue. In the computer-initiated interface, the computer selects the next stage in the interaction.

2. Computer Initiated Interfaces the computer is in charge, controlling the progress of the user/computer dialogue. Information is displayed and the user response of the computer takes action or displays further information.

Hardware Interfaces:
1) 1 GB of RAM and higher.
2) Processor Intel Pentium 4 (1.5 GHz) and above or equivalent.
3) 40 GB of Hard disc and higher.
4) Internet connection.

Software Interfaces:
1) Programming Language: Java
2) Tools: JDK 1.6 or above.
3) Operating System: Linux operating system

Methods:
Naïve Bayes Classifier
It's a probabilistic and supervised classifier given by Thomas Bayes. According to this theorem, if there are two events say, e1 and e2 then the conditional probability of occurrence of event e1 when e2 has already occurred is given by the following mathematical formula:

$$P(e1 \mid e2) = \frac{P(e2 \mid e1)\, P(e1)}{P(e2)}$$

This algorithm is implemented to calculate the probability of a data to be positive or negative. So, conditional probability of a sentiment is given as:

$$P(\text{Sentence} \mid \text{Sentiment}) = \frac{P(\text{Sentiment} \mid \text{Sentence})\, P(\text{Sentiment})}{P(\text{sentiment})}$$

And conditional probability of a word is given as:

$$P(\text{Sentiment} \mid \text{Word}) = \frac{\text{No of words occurrence in class} + 1}{\text{No of words belongs to class} + \text{Total no of words}}$$

Evaluation of Method:
To evaluate the method following measures are used:

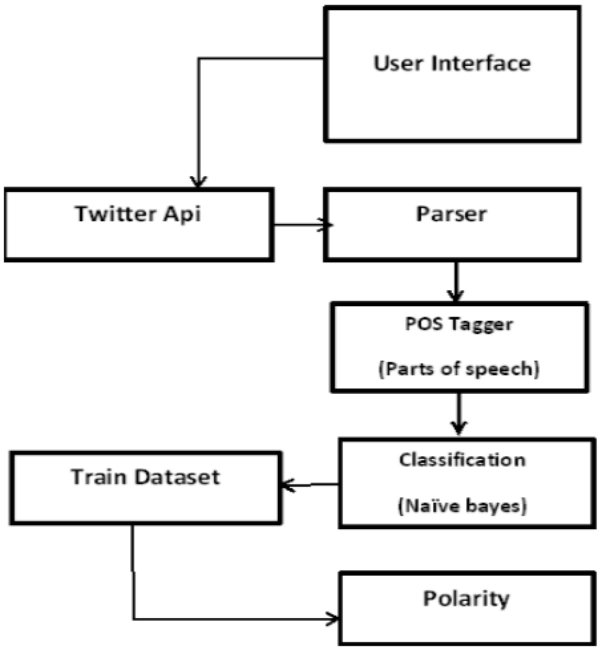
Accuracy, Precision, Recall, Relevance.

Following contingency table is used to calculate the various measures.

Relevant	Irrelevant
True Positive (tp)	False Positive (fp)
False Negative (fn)	True Negative (tn)

Precision = tp / tp + fp
Accuracy = tp + tn / tp + tn + fp + fn
Recall = tp / tp + fn

System Architecture:



1.User Interface: User interface is mainly used to taking input from user,it's a run-time input e.g. product name from user.

2. Twitter API: Based upon user input the information (reviews/tweets) will be collected from social media (twitter). For collection information from twitter JAVA-API of twitter will be used,whichbasically collect all the information up to the current instance of time. Twitter API provides functionality to getting tweets from twitter.

3.Parser: When the data will be collected from twitter API it will be raw/XML data,which will be hard to analysis. To convert raw XML data to something meaningful data i.e. extract only meaningful data from that whole data there will be a parser which will convert the XML data to meaningful data.

4. POS Tagger (Part of Speech Tagger): It's a java library made by Stanford University to parsing the sentence into part of speech.In opinion mining will be focusing on adjective,adverb,verb so to remove the unnecessary part of speech POS tagger will be used to filter out these things.

5. API to form Input: This will be bridge to convert the filtered POS in to the input form of classifier, which will be evaluated later using training data-set.

Training Data-set: The previous data history will be used to train the machine to differentiate the polarities on words.E.g. Good, positive Gd, positive bd, negative, not well, negative, :) –positive, :(-negative.

Based on training data-set the input will be processed.

Predict Output: The input will be processed based on trained data set for predicting the polarities on user reviews, e.g. positive or negativeness In the form of graphical.

Discussion:

In this paper, we focus on a challenging problem of predicting users' opinions toward topics they had not directly given yet, which we define as user-topic opinion prediction. The main contributions of this paper are as follows: 1) Different from previous work recognizing emotional states/sentiments from online micro-blogging data but ignoring whose they are, we seek to find out who has what opinion of a specific topic in advance. We believe that predicting individual's feeling about a given target is important for affective computing studies and able to be used to various applications. 2) To provide a solution, we consider the opinion among Twitter social friends and users' opinion consistency on content-related topics, and formulate them as social context and topical context mathematically. 3) Utilizing the learned emotional knowledge from the observed tweets and the social and topical context information, we propose a Naïve Bayes classification method to predict the unknown user-topic opinions.

REFERENCES:

1. Arun Monicka Raja M., Godfrey Winster S., Swamynathan S., "Review Analyzer: Analyzing Consumer Product Reviews from Review Collections," *IEEE International Conference on Recent Advances in Computing and Software Systems,* pp.287-292, Chennai, April 2012.
2. Kaigaun yu-Volume 56 No.13, October 2012, multiclass Svm,"Classi_cation of Opinion Mining Techniques", IEEE DOI 10.11.1109/IA.2014.7009461.
3. Ruyaa li, Amazon review,"METHODOLOGICAL STUDY OF OPINION MINING AND SEN-TIMENT ANALYSIS TECHNIQUES" IJSC vol-05, no-1, feb-2014.
4. Gang li-IJSC vol-56no-23, oct-12., K-means clustering,"opinion mining and analysis".
5. Gamgam-naive bayes,"Predicting User-Topic Opinions in Twitter with Social and Topical Context" IEEE DOI 10.1109/RACSS.2012.62126892.
6. Pravin Kumar singh, svm,"opinion mining and sentiment analysis"IEEE DOI 978-1-4799-6818-3/15/31:002015:
7. A. Go, R. Bhayani, and L. Huang, Twitter Sentiment Classi_cation Using Distant Supervision, technical Report, Stanford Univ., 2009.
8. B. Jansen, M. Zhang, K. Sobel, and A. Chowdury, Twitter Power: Tweets as Electronic Word of Mouth, J. Am. Soc. for Information Science and Technology, vol. 60, no. 11, pp. 2169-2188, 2009.
9. B. OConnor, R. Balasubramanyan, B. Routledge, and N. Smith, From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series, Proc. Intl AAAI Conf. Weblogs and Social Media
10. L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, Target-Dependent Twitter Sentiment Classi_cation, Proc. 49th Ann. Meeting of the Assoc. for Computational Linguistics: Human Language Technologies, vol. 1, pp. 151-160, 2011.