

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Paper / Case Study

Available online at: www.ijarcsms.com

Opinion Mining and Classification of User Reviews in Social Media

Gayathri Deepthi.V¹PG Scholar, Dept. Computer Science and Engineering
United Institute of Technology
Tamil Nadu, India.**K.Sashi Rekha²**Asst.Professor, Dept. Computer Science and Engineering
United Institute of Technology,
Tamil Nadu, India.

Abstract: Social media is increased in presence and importance in society. A social network service consists of a representation of each user. Social networking sites allow users to communicate with people in the network by sharing thoughts, pictures, status, posts, activities and products. It has become one of the biggest forums to express ones opinion. The majority of earlier work in Rating Prediction and Recommendation of products mainly takes the star ratings of users on products. However, most reviews are written in a free-text format which is difficult for computer systems to understand, analyze and aggregate. The proposed system is able to collect useful information from the social website and efficiently perform sentiment analysis of the reviews on product. The work focuses on identifying the sentiment information from free-form text reviews and using that information to rank the product. The sentiment of the user reviews is predicted using a well trained effective Naive Byes classifier. The result shows that using textual information given by the users is classified as positive negative and neutrals.

Keywords: Data Mining; Information Retrieval; Opinion Mining; Product Ranking; Sentiment Analysis;

I. INTRODUCTION

A social network is a collection of persons or organizations. The social relation could be both explicit (kinship and classmates) and implicit (friendship and common interest). The persons in the social network are considered as nodes. In this each node is connected with other node with number of links Social media has become one of the biggest forums to express ones opinion. Sentiment analysis is used to determine the attitude of a speaker or a writer with respect to some product. A basic task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or phrase level.

Sentiment analysis is the task of finding the opinion of the person. Sentiment analysis in sentence level is basically for finding weather the opinion is positive or negative sentiment. The analysis of digital texts can be performed using machine learning algorithm such as Naive byes.

Classifier, latent semantic analysis, support vector machines, and bag of words. When a person wants to buy a product online he/she will read the reviews written by other people on the various products. The sentences can be classified into two classes such as objective sentences and subjective sentences.

The Objective sentences contain factual information whereas subjective sentences contain clear opinions, beliefs and views about specific entities. These user reviews are a gold mine for companies and individuals that want to monitor their reputation and get timely feedback about their products and actions.

Sentiment analysis offers the people to choose the right product and also offers the organization to improve the quality of their product.

II. BACKGROUND AND RELATED WORK

Social media continues to gain increased presence and importance in society. Sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall contextual polarity of a document and the effect the user wants to have on the reader. Sentiment analysis has become popular in judging the opinion of consumers towards various brands [1]. The way in which consumers express their opinion on social networking websites helps to judge this opinion [2]. The main issue is to understand this sentiment and being able to classify it appropriately [3]. The tweets are obtained from the twitter website using the twitter API. This will provide us with a large source of information for conducting the sentiment analysis [4]. Since data is being retrieved from a micro blogging website an appropriate approach is to be used [5]. The tweets are first checked for relevance to Smartphone's by using a list of keywords [6]. The system is trained using the training dataset which makes it capable to analyze the input tweets [7]. The input tweets are then checked word by word and the words expressing opinion are taken into account [8]. The sentiment analysis of the tweets is performed by the system [9] after which the tweets are classified into positive, negative and neutral categories [10].

III. WORD EXTRACTION AND SENTIMENT ANALYSIS

To perform the sentiment analysis a trained dataset is considered. Data from the dataset is the input for the Entity Extraction module. The sentence will have some valuable information about its sentiment and the rest of the words will not give any clue regarding the sentiment. Such words should be removed by preprocessing. Data preprocessing is done to eliminate the incomplete, noisy and inconsistent data. Data must be preprocessed in order to perform any data mining functionality.

Data Preprocessing involves the following tasks Removing URLs, In general URLs does not contribute to analyze the sentiment in the informal text. For example consider the sentence "I have logged in to www.Ecstasy.com as I'm bored" actually the above sentence is negative but because of the presence of the word ecstasy it may become neutral and it's a false prediction. In order to avoid this sort of failures the URLs should be removed. After preprocessing the features are extracted. The Naïve Bayes Classifier is trained with a training data set labeled with sentiments positive, negative or neutral for sentiment analysis . There are one million labeled tweets in the data set. After performing preprocessing such as Stop Word removal, Stemming, and feature extraction on the input tweet, the Sentiment Classifier Model labels the tweet with a sentiment using this trained Naïve Bayes classifier.

IV. PROPOSED WORK

Social network has recently been increased in developing web relationships between individuals. Each Individual is allowed to give their reviews for a product in star rating and free form text. Textual information gives a better prediction than the star ratings given by the users. The sentiment information from the user's free-text reviews are identified and that knowledge is used for rating and ranking the product. User reviews are analyzed and classified at the sentence level as positive or negative. A Nobel Naive Bayes (NB) classifier [8] is used to classify the sentiment of the user reviews as positive or negative. This allows users to get recommendations on specific aspect of the product.

System Architecture Design

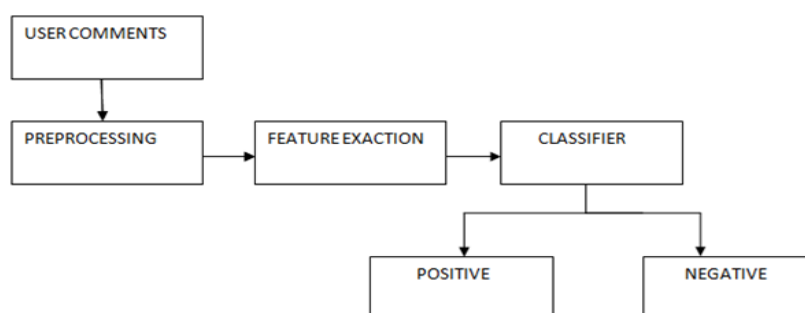


Fig 1 System Architecture Design

The system architecture of proposed system is shown in fig 1. All the users connected to the social network are allowed to enter their comments on specified products in. User normally enters their comments for the product in free text format. These comments are the reviews given by the user for a product. Reviews are then collected to rate the product.

Firstly, the collected reviews are pre-processed and the words are extracted. These extracted words are then stored in feature vector. The words are then classified into positive and negative. Based on the category of words the sentence is found as positive or negative sentence and the product is ranked.

V. EXPERIMENTAL ANALYSIS

Sentiment analysis where the tweets after being obtained from the twitter website are classified into positive, negative and neutral using the Naïve Bayes classifier. A Sample tweets is used which is then tested manually for accuracy. A threshold is set and the tweets are classified into positive, negative and neutral.

Matching Matrix (Confusion Matrix): A matching matrix is a specific table layout that allows visualization of the performance of an algorithm, typically an unsupervised learning one (in supervised learning it is usually called a confusion matrix). The instances in a predicted class are represented in each column of the matrix while the instances in an actual class are represented in each row.

TABLE 1
Sentiment Analysis Confusion Matrix

Predicted class Actual Class	Positive	Negative	Neutral
Positive	68	3	4
Negative	2	41	11
Neutral	5	7	69

Precision: Precision is a measure of the accuracy provided that a specific class has been predicted.

$$\text{Precision} = \frac{tp}{(tp + fp)}$$

where tp and fp are the numbers of true positive and false positive predictions for the considered class. The result is always between 0 and 1. In the matching matrix above, the precision for a class is calculated as:

- Positive = $68/(68+3+4) = 0.90$
- Negative = $41/(41+2+11) = 0.75$
- Neutral = $69/(69+11+1) = 0.85$

Accuracy: Accuracy is calculated as the sum of correct classifications divided by the total number of classifications. It is the overall correctness of the model.

$$\text{Accuracy for sentiment} = 168/200 = 0.840$$

Thus, the proposed system is able to collect useful information from the twitter website and efficiently perform sentiment analysis on the data using an efficient scoring system and a well trained Naïve Bayes Classifier, respectively

The sentiment analysis is performed for the sentence using a novel Naïve Bayesian Classifier. The sentiment of the user review is analyzed to know the attitude of the user and to know the product rank. The reviews are preprocessed for eliminating

noise and the words are extracted. The extracted words are classified into positive or negative using the Naïve Byessian Classifier. Thus, the proposed system is able to collect useful information from the twitter website and efficiently perform sentiment analysis on the data and predict the user's age and gender using an efficient scoring system and a well trained Naïve Bayes Classifier, respectively

VI. CONCLUSION

The sentiment analysis is performed for the sentence using a novel Naive Byessian Classifier. The sentiment of the user review is analyzed to know the attitude of the user and to know the product rank. The reviews are preprocessed for eliminating noise and the words are extracted. The extracted words are classified into positive or negative using the Naive Byessian Classifier. Thus, the proposed system is able to collect useful information from the twitter website and efficiently perform sentiment analysis on the data using a well trained Naive Byessian Classifier.

ACKNOWLEDGMENT

I am highly indebted to Associate Prof. Mr. M. Nageswara Gupta M.E., Head of the Department, Computer Science and Engineering for his encouragement towards the completion of this project.

With immense pleasure I would like to express my hearty thanks to my project guide, Asst.Prof. Ms.K.Sashi Rekha M.E., Department of Computer Science and Engineering for her encouragement and valuable guidance with keen interest towards the completion of this project.

References

1. Ali A. Ghorbani, Mostafa Karamibekr , "Verb Oriented Sentiment Classification" in International Conferences on Web Intelligence and Intelligent Agent Technolog 2012IEEE .
2. Bhaskar Prasad Rimal, Eunmi Choi, Ian lumb, "A Taxonomy and Survey of Cloud Computing Systems", Fifth International Joint S.Chan,C. Khoo J.C. Na, H. Sui and Y.Zhou.Effectiveness of simple linguistic processing in automatic sentiment classification of product reviews,Advances In Knowledge and organization, 2004, pages 49-54.
3. R. Prabowo and M. Thelwall, Sentiment analysis: A combined approach, Journal of Infometrics, 2009, pages 143-157.
4. Barla Cambazoglu, Hakan Ferhatosmanoglu, Ingmar Weber and, Hakan Ferhatos manoglu A large-scale analysis for Yahoo! Answers, Fifth ACM International Conference on Web Search and Data Mining, Seattle, Washington, USA, 2012, pages 633-642.
5. Adam Bermingham and Alan F. Smeaton, Classifying Sentiment in Microblogs:is brevity an advantage? Nineteenth ACM International Conference on Information And knowledge management, Toronto, Canada, pages 1833-1836.
6. Ana-Maria Popescu, Marco Pennacchiotti, Detecting Controversial Events from Twitter, Proceedings of the 19th ACM international conference on Information and knowledge management, Toronto, ON, Canada2010, DOI:0.1145/1871437.1871751, pages 1873-1876.
7. L.Lee, and Pang Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval, 2008, pages 1-135.
8. AleksandrVoskoboynik Eugene Agichtein, Jeff Pavel, Luis Gravano, ViktoriySokolova, and, Snowball:a prototype system for extracting relations from large text Collections Proceedings of the fifth ACM conference on Digital Libraries, Bremen.
9. ApoorvAgarwal, Boyi Xie, Ilia Vovsha, Owen Rambow and Rebecca Passonneau , Sentiment Analysis of Twitter Data, Proceedings of the Workshop on Languages Social Media in, Portland, Oregon, 2011, pages 30-38.
10. JunlanFeng, Luciano Barbosa, Robust sentiment detection on twitter from bayes and noisy data, Proceedings of the 23rd international Conference on Computational Lisguistics, Beijing,China 2010.

AUTHOR(S) PROFILE



GAYATHRI DEEPTHI V received B.TECH degree in Information Technology from Annai Mathammal Sheela Engineering College, Anna University in 2010. Presently she is pursuing her Master of Engineering in Computer Science and Engineering at United Institute of Technology, Tamil Nadu, India.



K.Sashi Rekha received M.E. degree in Computer Science and Engineering from Anna University Trichy. Presently she is working as Assistant Professor in United Institute of Technology, Tamil Nadu, India.