

Textual Reviews Based on Sentiment Analysis using Natural Language Processing

Neha mane C¹, Rrakshitha M², Shilpa Awari³, Mohammed Sarvar⁴
^{1,2,3,4}Vivekananda Institute of Technology, Bangalore, India.

Abstract- Sentiment analysis is an application of natural language processing. We are focusing on textual analysis. This is a very popular field of research in text mining. It is also known as emotion extraction or opinion mining. We basically combine user opinions to convert into measured rating system for better understanding of the response given by the user. The basic idea is to find the polarity of the text and classify it into positive, negative or neutral. It helps in review decision making. This paper presents the survey of main approaches used for sentiment classification.

Keywords- text mining; rating system; sentiment classification; feature extraction; polarity

I. INTRODUCTION

Sentiment analysis is the process of extracting emotions or opinions from a piece of text for a given topic. It allow us to understand the attitudes, opinions and emotions in the text. In this, user's likes and dislikes are captured from web content. It involves predicting or analyzing the hidden information present in the text. This hidden information is very useful to get insights of user's likes and dislikes. The aim of sentiment analysis is to determine the attitudes of a writer or a speaker for a given topic. Sentiment analysis can also be applied to audio, images and videos. Today internet has become the major part of our life. Most of the people use online blogging sites or social networking sites to express their opinions on certain things. They also use these sites to know what other people's opinions are. Thus mining of this data and sentiment extraction has become an important field of research. For example, the customer will decide what to buy if he or she sees valuable reviews posted by others, especially user's trusted friend. We believe reviews and reviewers will do help to the rating prediction based on the idea that high-star ratings may greatly be attached with good reviews.

Hence, how to mine reviews and the relation between reviewers in social networks has become an important issue in web mining, machine learning and natural language processing. In many practical cases, it is more important to provide numerical scores rather than binary decisions. Generally, reviews are divided into two groups, positive and negative. However, it is difficult for customers to make a choice when all candidate products reflect positive sentiment or negative sentiment. To make a purchase decision, customers not only need to know whether the product is good, but also need to know how good the product is. It is also agreed that different people may have different sentimental expression preferences. For example, some users prefer to use "good" to describe an "excellent"

product, while others may prefer to use "good" to describe a "just so" product.

In our daily life, customers are most likely to buy those products with highly-praised reviews. That is, customers are more concerned about item's reputation, which reflects consumers' comprehensive evaluation based on the intrinsic value of a specific product. To obtain the reputation of a product, sentiment in reviews is necessary. Normally, if item's reviews reflect positive sentiment, the item may be with good reputation to a great extent. Oppositely, if item's reviews are full of negative sentiment, then the item is to be with bad reputation. To a given product, if we know user sentiment, we can infer the reputation and even the comprehensive ratings. When we search the net for purchasing, both positive reviews and negative reviews are valuable to be as reference. For positive reviews, we can know the advantages of a product. For negative reviews, we can obtain the shortcomings in case of being cheated.

II. SYSTEM FRAMEWORK AND CONCEPTS

The proposed system in our paper utilizes all the existing concepts which have been combined to provide better clarity in the working and the result given. Any text or statements are broken into the sets of the following assessments in order to implement them in the algorithms used.

A. Subjectivity/Objectivity

To perform sentiment analysis we first need to identify the subjective and objective text. Only subjective text holds the sentiments. Objective text contains only factual information. Example:1.Subjective: Interstellar is an epic film. (this sentence has a subjective word. (epic) 2.Objective: Zack Snyder is the director of 300. (this sentence is a fact without any sentiment)

B. Polarity

Further subjective text can be classified into 3 categories based on the sentiments conveyed in the text.

- Positive: I really appreciate the service provided.
- Negative: The food at that restaurant was awful.
- Neutral: I get sleepy by 12. This positive, negative and neutral nature of text is termed as polarity of text.

C. Sentiment level

Sentiment analysis can be performed at various levels.

- Document Level: The whole document is given a single polarity positive, negative or objective.
- Sentence Level: In this, a document is classified at sentence level. Each sentence is analyzed separately and classified as negative, positive or objective. Thus

overall document has a number of sentences where each sentence has its own polarity.

- **Phrase Level:** It involves much deeper analysis of the text and deals with identification of the phrases or aspects in a sentence and analyzing the phrases and classify them as positive, negative or objective. It is also called aspect based analysis.

III. APPLICATIONS

A. *Support in decision making*

Decision making is a very important field of our life. Opinions extracted from reviews helps us in making various decisions like “which cosmetic to buy”, “which cafe to go”, “which play to watch” etc.

B. *Business application*

In today’s world of competition, every company wants to satisfy its customer’s requirements by creating innovative products. Assessments of individuals are an essential angle today with the goal that organizations can get an input from clients and can roll out sought improvements in their item. Google Product Search is one illustration.

C. *Predictions and trend analysis*

Sentiment analysis enables one to predict market trends by tracking views of public. It is also helpful in elections where candidate wants to know the expectations of people from them.

D. *On social media*

Few blogs, and online forums that millions of people are busily discussing and reviewing businesses, companies, and organizations. And those opinions are being ‘listened to’ and analyzed. The results from sentiment analysis help businesses understand the conversations and discussions taking place about them, and help them react and take actions accordingly. They can identify the negative reviews and causes and turn them into good reviews by formulating better strategies and plans all in turn providing with better returns in investments and efforts.

IV. METHODS USED

A. *Support Vector Machine*

In general, Support Vector Machine (SVM) is a two class classification model. The definition of the basic model is a linear classifier with interval maximum in feature space. Its learning strategy is the maximum distance, can be transformed into solving a convex quadratic programming. Text data is ideally suited for SVM classification because of the sparse nature of text, in which few features are irrelevant, but they tend to be correlated with one another and generally organized into linearly separable categories. The feature space is large for any text classification task; therefore, linear kernel is usually used in rating prediction performances.

B. *Naive Bayesian Model*

Naive Bayes method is a classification method based on Bayesian theorem and assumption of independent

characteristic conditions. The Naive Bayes classifier is based on a simple assumption that the property is independent of each other when given target value.

C. *Maximum Entropy*

The principle of maximum entropy (ME) means that when the probability distribution of a random event is predicted, the prediction should satisfy all the known constraints, and do not make any subjective assumptions about the unknown. In this case, the probability distribution is the most uniform, and the minimum risk is predicted. ME does not have to rely on the assumption of independent feature items, and it can treat all possibilities equally. Without regarding to the interaction and influence between the various corpus, it can be an arbitrary to increase the training corpus. The richer the training set, the more improved the accuracy of model study.

D. *Artificial Neural Network*

Artificial Neural Network is able to handle real data and its structure symbolizes human brain and neural system, so that this method is a powerful implementation tool for prediction. Artificial Neural network relies on the idea of derive features from linear combinations of the data provided as an input, and model the output as a nonlinear function of these features. ANN contains a variety of methods, such as BP (Back Propagation) neural network, RBF (Radial basis function) neural network, feedback neural network and Deep Learning.

V. METHODOLOGY

Feature selection: To perform sentiment classification, first task is to extract the features from text which are:

A. *N grams*

n grams refers to consecutive n terms in text. One can take only one word at a time(unigram) or two words(bigram) up to n accordingly. Some sentiments can’t be captured with unigram feature. For example, this drink will knock your socks off. It is a positive comment if socks off is taken together and negative in case of only unigram model (off).

B. *POS tagging*

It is a way towards denoting a word in a content(corpus) as comparing to parts of speech in light of both its definition and its association with contiguous words. Nouns, pronouns, adjectives, adverbs etc are the parts of speech. Adjectives and adverbs hold most of the sentiments in the text.

C. *Stemming*

It is the process of removing prefixes and suffixes. For example, ‘playing’, ‘played’ can be stemmed to ‘play’. It helps in classification but sometimes leads to decrease in classification accuracy.

D. *Stop words*

Pronouns (he/she, it), articles (a, the), prepositions (in, near, beside) are stop words. They provide no or little information about sentiments. There is a list of stop words available on

the internet. It can be used to remove them in the pre-processing step.

E. Conjunction handling

In general, each sentence expresses only one meaning at a time. But certain conjunction words like but, while, although, however changes the whole meaning of the sentence. For example, although movie was good but it was not up to my expectations. By using these rules throughput can be increased by 5%.

F. Negation handling

Negation words like 'not' inverts the meaning of whole sentence. For example, the movie was not good has 'good' in it which is positive but 'not' inverts the polarity to negative.

VI. CONCLUSION

In this paper, a recommendation model is proposed by mining sentiment information from social users' reviews. We combine user sentiment similarity, interpersonal sentiment influence, and item reputation similarity into a unified matrix factorization frame-work to achieve the rating prediction task. In particular, we use social users' sentiment to denote user preferences. Besides, we build a new relationship named interpersonal sentiment influence between the user and friends, which reflects how users' friends influence users in a sentimental angle. What is more, as long as we obtain user's textual reviews, we can quantitatively measure user's sentiment, and we leverage items' sentiment distribution among users to infer item's reputation. The experiment results demonstrate that the three sentimental factors make great contributions to the rating prediction. Also, it shows significant improvements over existing approaches on a real-world dataset. In our future work, we can consider more linguistic rules when analyzing the context, and we can enrich the sentiment dictionaries to apply fine-grained sentiment analysis. Besides, we can adapt or develop other hybrid factorization models such as tensor factorization or deep learning technique to integrate phrase-level sentiment analysis. In recent years, the information on the social network has exploded, and the sentiment classification has become more and more important. In this paper, the machine learning methods in SA in recent years are summarized, and the formulas of traditional methods (such as SVM, NB and ME) are given in detail. We also look forward to the latest ANN (BPN and CNN) methods. Finally, the affective methods and challenges of emotion analysis are given. As a mature research point, SA becomes more and more challenging because of the rise of deep learning. Sentiment Analysis is a technique that determines whether the attitude of a person is positive, negative or neutral based on their writing. The system can withstand huge volumes of customer reviews and provides better accuracy.

VII. REFERENCES

- [1]. A Survey of Sentiment Analysis Techniques – (I-SMAC 2017).
- [2]. Sentiment Analysis for Restaurant Rating – 2017 IEEE International Conference on Smart technologies and management.
- [3]. Sentiment Analysis for Reviews of Restaurants in Myanmar Text.
- [4]. A Survey on Sentiment Analysis by using Machine Learning Methods- Peng Yang.

TEXT BOOK REFERENCES

- [1] Hands on Machine Learning with Scikit Learn and Tensor Flow first edition.
- [2] The Elements of Statistical Learning fourth edition.
- [3] Introduction to Machine Learning with Python first edition.
- [5]. Pattern Recognition and Machine Learning third edition.



GUIDE: NIBEDITA P

Assistant Professor Dept. of Computer Science Engineering
Bachelor of Engineering and M.Tech in Biju Patnaik
University of Technology



AUTHOR 1: NEHA MANE C

Dept. of Computer Science Engineering Bachelor of
Engineering in Visvesvaraya Technological University



AUTHOR 2: RAKSHITHA M

Dept. of Computer Science Engineering Bachelor of
Engineering in Visvesvaraya Technological University



AUTHOR 3: SHILPA AWARI

Dept. of Computer Science Engineering Bachelor of Engineering in Visvesvaraya Technological University



AUTHOR 4: MOHAMMED SARVAR

Dept. of Computer Science Engineering Bachelor of Engineering in Visvesvaraya Technological University