

Product Quality Evaluation with the help of customer reviews by using different Data mining technique

¹ Mayuri Saste, ² Darshana Jinde, ³ Sanjokta Malusare, ⁴ Archana Kapare, ⁵ Prof. M. A. Rane

Abstract:

In the modern world, online shopping trend is increasing day by day. People share their opinions, feedbacks and experiences in the form of reviews. This user generated content is very useful for business, government and individuals in decision making. Processing and analyzing this huge opinionated information is difficult and time consuming. Hence there is a need of intelligent system which automatically mines and classifies this data into positive, negative and neutral categories. Sentiment Analysis is a technique in which sentiment is gathered, analyzed and aggregated from text. Sentiment analysis is used for mining this data but to obtain more fine grained analysis we need to go to Aspect Based Sentiment Analysis (ABSA). Data mining, natural language processing and statistical approaches are utilized based on dictionaries, grammatical analysis and semantic understanding of text. Dependency relations use grammatical representation of sentence structure as a set of relationships among entities. ABSA is decomposed into three main subtasks: aspect term extraction, aspect term aggregation and aspect term rating.

Index Terms - Component, formatting, style, styling, insert.

INTRODUCTION

The growth of social media contributes huge amount of data content such as comments, reviews and opinions about product, services, etc. This data is based on sentiments which are essential for customers as well as for manufacturers. So there is need of system which can analyse and summarize opinions for a particular aspect of a product. For example, consider “The display of phone is perfect”, “display” is aspect for which opinion is expressed. In this paper, we propose an unsupervised model which makes use of dependency relations to detect aspects and opinions from reviews in sentiment analysis system (SAS). Existing approaches fail to cover the fact that if two reviews are mentioning same aspect with two different words they are considered as two different aspects. It also fails to address the problem such as recently updated reviews are shown first which might conflict with old reviews. This misleads customers while making decision. The system also provides aspect wise rating which helps customer to make better decision while buying the product. The proposed system also removes conflicts between reviews and ratings. It also acts as a recommendation system.

Keyword: Natural Language Processing, Abstract Based Sentiment Analysis, Data mining, Dependency relations, Machine Learning.

LITERATURE SURVEY

It is found that a system that can identify aspects accurately from given text corpus is developed. It is a system that can provide aspect wise results in the order of quality of aspects when searched in order to achieve satisfying results. Natural Language Processing Algorithm used this project. There is a need to inject external knowledge to improve the results. As data is unbalanced and in different forms there is need to explore machine learning techniques that address this problem [1].

Chien-Liang Liu and Wen-Hoar Hsaio aimed to design and develop a movie-rating and review-summarization system in a mobile environment. The movie-rating information is based on the sentiment-classification result. They have proposed a novel approach based on latent semantic analysis (LSA) to identify product features. The design is limited for movie domain only. The system is similar but not applicable for e-commerce domain [2].

The goal is to identify opinions expressed about specific entities (e.g., laptops) and their aspects (e.g. price). The techniques used in this paper are NLP Algorithm as well as SVM (Support Vector Machine) for Classification. In this paper Sentiment Analysis technique is used for finding different Aspect [4]. Future work includes applying the new framework and annotation schema to other languages (e.g. Spanish, Greek). Basically Sentiment Analysis does not include other language like as Spanish, Greek. [3]

Many recently proposed algorithms enhancements and various Sentiment Analysis applications are investigated and presented briefly in this survey. The main contributions of this survey include the sophisticated categorizations of a large number of recent articles and the illustration of the recent trend of research in the sentiment analysis and its related areas. Using NLP tools to reinforce the Sentiment Analysis process has attracted Researchers recently and still need some enhancements. [4]

Sagar Bhuta, Uehit Doshi, Avit Doshi and Meera Narvekar described a number of techniques, both lexicon-based approaches as well as learning based methods that can be used for sentiment analysis of text. Authors also highlights a number of issues and challenges that need to be overcome for sentiment analysis of Twitter data. The issue with the lexicon based approach is lexicons are not available in all languages. Also because of large amount of twitter data and usage of slang and casual use of language makes analyzing sentiment difficult [5].

Mira Dholariya, Dr.Amit Ganatra and Prof. Dhaval Bhoi described the survey on main approaches for performing sentiment analysis, Different tools for Sentiment Analysis, Application Area of Sentiment Analysis. The authors have discussed Sentiment Classification (SC) techniques such as machine learning approach uses supervised and unsupervised learning and Lexicon-construct approach classified into dictionary-based approach and corpus-based approach. This survey gives Illustration of the recent trend of research in the sentiment analysis and its related areas [6].

MATHEMATICAL MODEL

Mathematical model set theory $S = \{s, e, X, Y, \Phi\}$

s= Start of the program

1. Register/Login into the system
2. Provide comment and review of separate category.

e= End of the program

Identify the Product Rating

$X = \text{input of the program} = \{P, R, Q\}$

P = Review

R= Answer

Q=Rating of product

Y= Output of program= best product Rating

First, users provide feedback for specific rating out offs (1-5).

Let R be the set of rating

$R=\{R1, R2, R3... Rn\}$

Let A be the set of categories

therefore,

$A=\{A1, A2, A3..., Am\}$

$E= \{E1, E2, E3,..., Em\}$

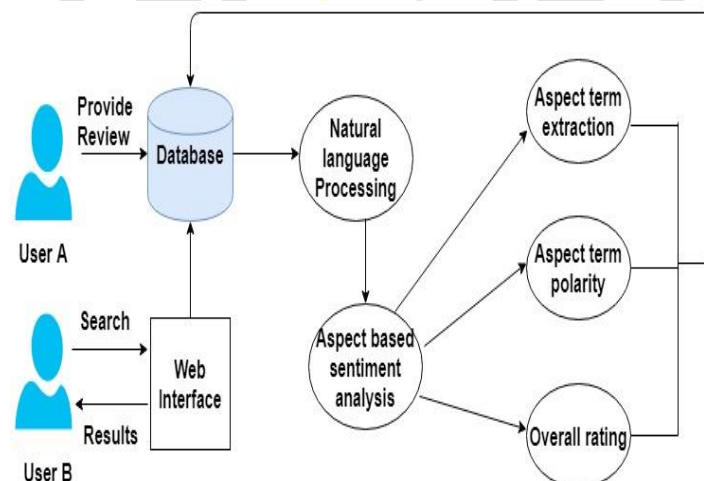
Overall rating is evaluated with the help of these ratings which basically represents quality of the product.

$Y= E1+E2+... +Em / m$

Where m is number of overall rating.

SYSTEM ARCHITECTURE

Our system is e-commerce based application which consists of electronic products like mobile, television and laptops. The user will provide feedback for the products. Our system will identify the aspects from the reviews along with its sentiments by using aspect based sentiment analysis which will give better search results. In this we have provided the ratings to the specific aspects as per the user reviews and using this reviews the quality of the product is evaluated .



METHODOLOGY

Subtask 1: Data pre-processing

Stanford Core NLP tools are used for preprocessing: lemmatization, split sentence, POS tagging and dependency parsing. □

- Data Cleaning: Removing non-relevant data or symbols from reviews.
- Lemmatization: Once the review is cleaned, lemmatization is done. It reduces the word into dictionary form. Let's take an example, "Awesome phone. I loved the UI interface and the way it looks. It is fast and smooth. Camera works well in both low light and normal light. Battery is also good." Here words "loved", "looks", "works" become "love", "look", "work" that is its base form.
- Split sentence: Splitting the whole review into sentence allows the extraction at sentence level. The opinion and aspect can be found within the sentence.
- POS tagging: The aim of POS tagging to assign part of speech to each word such as noun, verb, adverb, etc. For example camera battery are nouns in above example.
- Dependency parsing: In this a syntactic structure is assigned to the sentence and the relationships between head words and words is determined.

Subtask 2: Aspect term and opinion extraction

The subtask involves aspect and opinion extraction, which are simultaneously executed. However, as per research some assumptions such as nouns are aspects and adjectives are opinions is made.

Subtask 3: Rating individual aspect by taking average

Once we get aspect and its associated opinion, opinion word is mapped with opinion lexicon dictionary and rating is given from scale 1 to 5 for the particular aspect. At last average rating of all the aspects is displayed.

CONCLUSION

Sentiment analysis is the study that is used to analyze people feedbacks, sentiments towards the product. This paper is used to perform evaluation measure on feedbacks obtained from the customer. This paper removes the conflicts between reviews and ratings. In this paper, we have used Natural language processing technique in order to perform sentiment analysis of aspects. In order to retrieve explicit features and their sentiments, the dependency parsing algorithm is used. After classifying the positive and negative words by using aspect based sentiment analysis, the overall polarity is calculated. Thus the positive and negative separation of reviews is used to analyze the quality of the online products.

ACKNOWLEDGMENT

It gives us great pleasure in presenting the preliminary project report on **'Product Quality Evaluation with the help of customer reviews by using different Data mining technique**

We would like to take this opportunity to thank my internal guide Prof. Mugdha Rane for giving us all the help and guidance we needed we are really grateful to them for their kind support. Their valuable suggestions were very helpful.

We are also grateful to our Head of Information Technology Department, for her indispensable support and suggestions.

Name of Students

¹ Mayuri Saste, ² Darshana Jinde, ³ Sanjokta Malusare, ⁴ Archana Kapare

REFERENCES

- [1] Kim Schouten, Onne van der Weijde, Flavius Frasincar, and Rommert Dekke., “Supervised and Unsupervised Aspect Category Detection for Sentiment Analysis with Co-Occurrence Data,” IEEE Trans. Syst., 2017.
- [2] Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu, and Emery Jou, “Movie Rating and Review Summarization in Mobile Environment,” IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., Vol. 42, No. 3, May 2012.
- [3] Maria Pontiki, Dimitrios Galanis¹, Haris Papageorgiou¹, Ion Androutsopoulos, “SemEval-2016 Task 5: Aspect Based Sentiment Analysis,” 2016.
- [4] Walaa Medhat a, Ahmed Hassan b, Hoda Korashy, “Sentiment analysis algorithms and applications: A survey,” 2014.
- [5] Sagar Bhuta, Uehit Doshi, AvitDoshi, Meera Narvekar, “A Review of Techniques for Sentiment Analysis of Twitter Data,” 2014.
- [6] Mira Dholariya, Dr. Amit Ganatra, Prof. DhavalBhoi, “A Surveyon Sentiment Analysis: Tools and Techniques,” Vol. 5, Issue 3, March 2017.

