PRODUCT REVIEW SENTIMENT ANALYSIS WITH TRENDY SYNTHETIC DICTIONARY AND AMBIGUITY MITIGATION

Prof. A. N. Nawathe
anunawathe@gmail.com

Godge Isha Sudhir

ishagodge37@gmail.com

Dang Poornima Mahesh

dang.poornima@gmail.com

Arvikar Shruti Sanjay

shrutiarvikar89@gmail.com

Maske Rushikesh Kantrao

rushimaske326@gmail.com

Abstract

Data mining is the process of turning raw data into useful information. The main use of data mining is to fetch the required data and extract useful information from the data and to interpret the data. In the existing system, Bag of Words model is used along with Dual sentiment Analysis in order to classify the reviews as positive, negative and neutral. However, the performance of Bag of Words sometimes remains limited due to some fundamental deficiencies in handling the polarity shift problem. The proposed system uses a dictionary based classification for accurately classifying the reviews as positive, negative and neutral. The proposed system additionally analyses the flaws of the existing systems and thereby propose two major features such as identifying the negation oriented sentiments and the conjunction oriented sentiments which require the analysis of pre-conjunction and post conjunction sentences. So the ambiguity is reduced by analyzing such conjunction and negation based sentences. To enhance the accuracy in the classification of neutral reviews, Dual sentiment analysis method is implemented. Both the product owner and the user can identify the quality of the product based on the sentiment graph that is generated based on the reviews for each of the product.

Keywords*:Dual Sentiment Analysis, POS Tagging, Sentiment Polarity, Tokenization, Opinion Mining, Data Mining.*

1. INTRODUCTION

Currently with the growing amount of online reviews available on the Internet, sentiment analysis and opinion mining, as a special text mining task for determining the subjective attitude (i.e., sentiment) expressed by the text, is becoming a hotspot in the field of data mining and natural language processing. Sentiment classification is a basic task in sentiment analysis, with its object to classify the sentiment (e.g., positive or negative) of a given text. Analyzing these reviews will enhance both the end users and the account executive. Sentiment analysis has earned its identification and is used in classifying the reviews. Sentiment analysis, also known as opinion mining, is the field of study that analyses people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes. It represents a large problem space also there are more names and slightly different tasks, e.g., opinion mining, opinion extraction, sentiment analysis, sentiment mining, subjectivity analysis, affect analysis, emotion analysis, review mining, etc. However, they are now all under the umbrella of sentiment analysis or opinion mining.

In academia both sentiment analysis and opinion mining are frequently employed. They basically represent the same field of study. The meaning of opinion itself is still very broad. Sentiment analysis and opinion mining mainly focuses on opinions which express or imply positive or negative sentiments. To do an analysis, classification plays a key role in opinion mining. A Classification Algorithm is a procedure for selecting a hypothesis from a set of alternatives that best fits a set of observations. Opinions are central to almost all human activities because they are key influencers of our behaviors. Whenever there is a need to make a decision, others' opinions are required. In the real world, businesses and organizations always want to find consumer or public opinions about their products and services. In previous, when an individual needed opinions, she/he asked friends and

family. When a business or an organization required public or consumer opinions, it conducted surveys, opinion polls, and focus groups. Acquiring consumer and public opinions has long been a huge business itself for marketing, public relations, and political campaign companies.

Opinion summarization describes opinions of articles by telling sentiment polarities, correlated events and the degree and with opinion summarization, a customer can easily view how the existing customers feel about a product, and the product manufacturer can get the reason why different stands people like it or what they complain about. A seller's job can be quite complicated or it can be quite easy. The two contradictory terms define the selling experience, based on the fact as how seller interprets the consumer interests. Unless one is a psychic or knows how to get into others mind the actual demand of the customer's and the product can't be collaborated. Having a right product is important and equally important is to present it before the right customer.

2. LITERATURE SURVEY

2.1 " Sentiment Analysis and Opinion Mining"

This concept is the field of study that analyzes sentiments, peoples' opinions, evaluations, attitudes, and emotions from written language. This technique is most active research areas in natural language processing and is also widely studied in data mining, Web mining, and text mining. This research has scope outside of management sciences to the computer science and social sciences due to its importance to business and society as a whole. The gaining importance of sentiment analysis coincides with the develop of social media like as reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks. We have a large opinionated data recorded in digital form for analysis. This systems are being applied in every business and social domain because opinions are central to almost all human activities and are key influencers of our behaviors. Our perceptions and beliefs of reality, and the choices we make, are largely conditioned on how others see and evaluate the world and when we need to make a decision we often seek out the opinions of others. This is true not only for individuals but also for organizations.

2.2 "Sentiment Analysis Model for Polarity Classification Based On Movie Reviews Using Lexicon Based Technique"

Bag-of-words (BOW) is now the most popular way to model text in statistical machine learning approaches in sentiment analysis. However, performance of Bag of Words sometimes remains limited due to few fundamental deficiencies in handling the polarity shift problem. The propose model called dual sentiment analysis (DSA) address the problem for sentiment classification. Firstly propose a novel data expansion technique by creating a sentiment reversed review for each training and test review. On this base, propose a dual training algorithm to make use of original and reversed training reviews in pairs for learning a sentiment classifier, and a dual prediction algorithm to classify and check the test reviews by considering two sides of one review. It extend the DSA framework from polarity (positive-negative) classification to 3-class (positive negative- neutral) classification, by taking the neutral reviews into consideration and by the dual sentiment analysis propose approach is to analysis sentiment as well as its automatic rating count. This can be calculated by using user review on the basis of positive, negative and neutral response. Then calculate all review and display the result analysis.
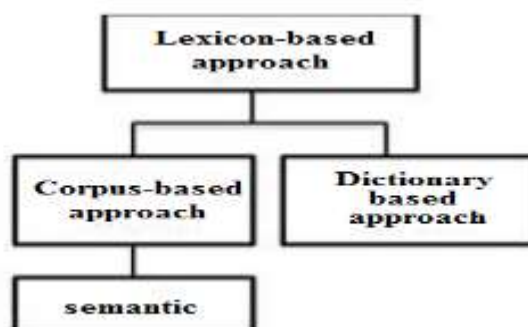

Fig 2.2: Lexicon based approach

2.3. "Sentiment Analysis : On Product Review"

The Web technology has rapidly changed the way that people express their views and opinions about any product. Now if one wants to purchase a product, people are no longer limited to asking their friends and families because there are many product

reviews on the Web which give opinions of existing users of the product. Here we show the system which facilitates us information about such products and services in summarization form. Finding opinion sources and monitoring them on Web can still be a difficult task because there are large numbers of different sources, and each source may have a huge volume of text with opinions or sentiments. In major cases, opinions are hidden in conference posts and blogs. It is complicated for a human reader to find relevant sources, extract related sentences with suggestions, read them, summarize them, and manage into usable forms. Thus, automated summarization systems are needed. Using this summarization we can identify the importance, quality, popularity of product and services. In this system we can make summarization for product but we can use this system anywhere, where text analysis is required. Sentiment analysis is also known as opinion mining, grows out of this need. It is challenging natural language processing or text mining problem. Due to its tremendous value for practical applications, there has been an excessive growth of both research in academia and applications in industry.
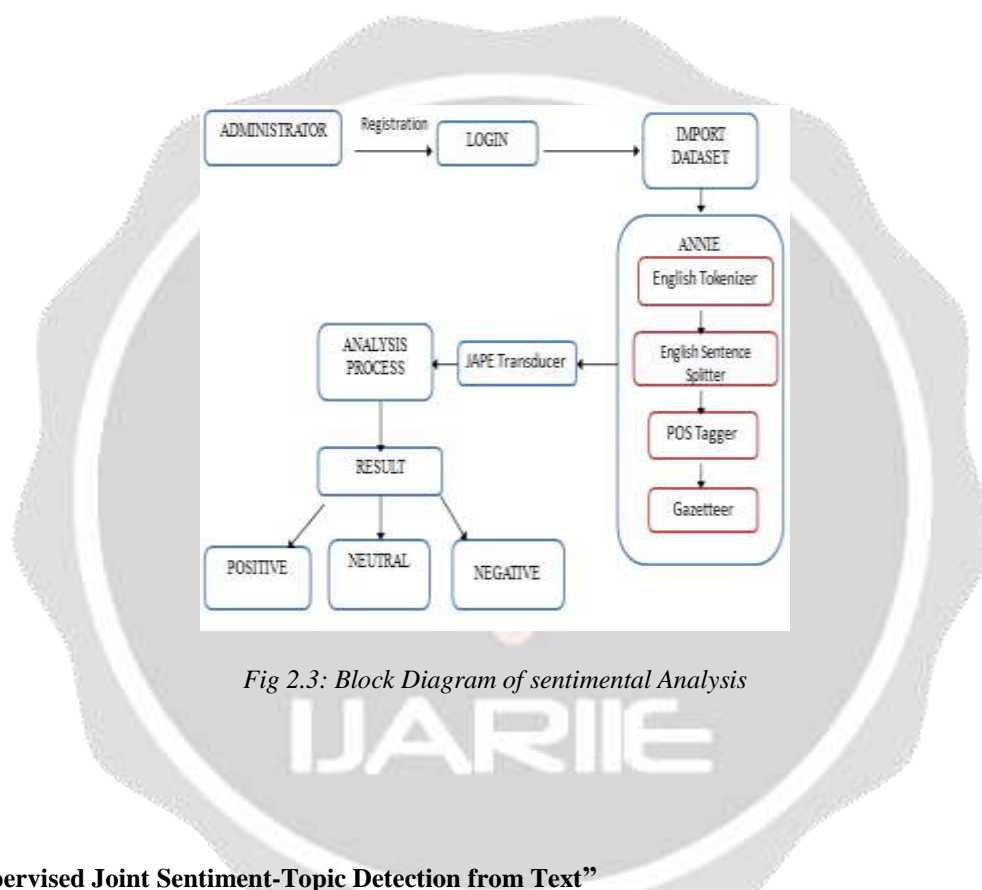

Fig 2.3: Block Diagram of sentimental Analysis

2.4 "Weakly Supervised Joint Sentiment-Topic Detection from Text"

Opinion mining or sentiment analysis aims to use automated tools to detect subjective information such as opinions, attitudes, and feelings expressed in text. This technique proposes a novel probabilistic modelling framework called joint sentiment-topic (JST) based on latent Dirichlet allocation (LDA), which detects sentiment and topic simultaneously from text. Joint sentiment topic model called Reverse joint sentiment-topic, obtained by reversing the sequence of topic generation and sentiment in the modeling process, is also studied. Results presents that when sentiment priors are added then JST performs better than Reverse-JST. This is inspected by the practice results on data sets from five different domains. Topics and topic sentiment detected by JST are indeed informative and coherent.

2.5 "Dual Sentiment Analysis: Considering Two Sides of One Review"

Data mining is the process of turning raw data into useful information. The main use of data mining is to fetch the required data and extract useful information from the data and to interpret the data. In the existing system, Bag of Words model is used along with Dual sentiment Analysis in order to classify the reviews as positive, negative and neutral. However, the performance of Bag of Words sometimes remains limited due to some fundamental deficiencies in handling the polarity shift problem. The proposed system uses a dictionary based classification for accurately classifying the reviews as positive, negative and neutral. The proposed

system additionally analyses the flaws of the existing systems and thereby propose two major features. To enhance the accuracy in the classification of neutral reviews, Dual sentiment analysis method is implemented. Both the product owner and the user can identify the quality of the product based on the sentiment graph that is generated based on the reviews for each of the product.

3. PROPOSED SYSTEM

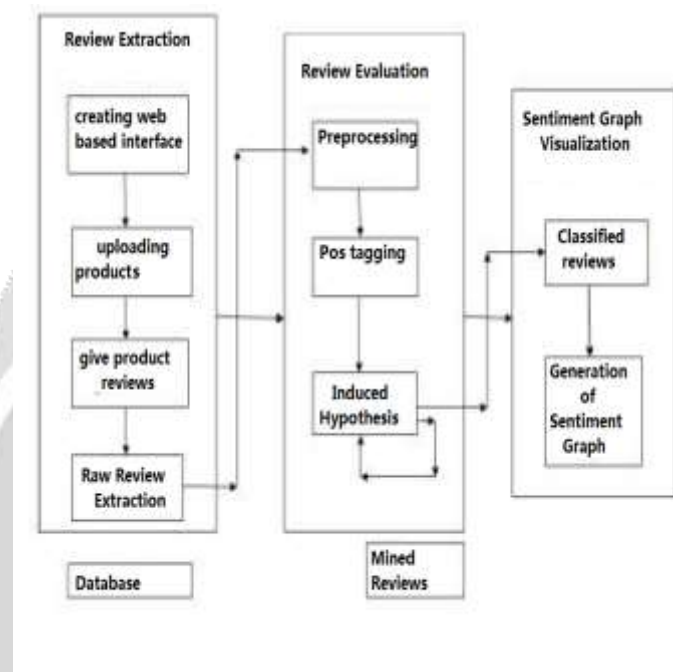### 3.1 System Architecture

Fig 3.1: Architecture of Proposed System

The proposed system consists of three major steps:

1. Reviews Extraction

- i. Creation of user interface and uploading images

- ii. Word clustering

- iii. Pre-processing

2. Reviews Evaluation

- i. Selection of hypothesis set

- ii. Classification

3. Sentiment graph visualization.

The first step of the module involves creation of web-based interface. Upon creation, the admin will login and upload images. Then, the user will login and view images. He shall then give reviews for the image which will get stored in the database. The stored reviews are then processed with the help of word clustering and pre- processing. In review evaluation, refining of the text

in the form of relevant phrases, words with the help of classification is performed. Classification is based on the supervised learning. The algorithm classifies the words given by the user into positive, negative and neutral. The words that do not have any meaning is considered as a hypothetical word. Such kind of words is processed with the help of an iterative induced hypothesis. The last module involves generation of a sentiment graph. It takes the classified reviews that are stored in the database as a input which are visually shown as the sentiment graph.

3.2. MODULES of PROPOSED SYSTEM

There are five modules in the proposed system, which are as follows:-

a.   Extraction and Pre-processing

In this module review of user's feedback is formulated, extracted and pre-processing. Users are those who have valuable in-put and feedbacks. Users who are more familiar with informative sites and can use our features efficiently. These valuable feeds will lead to enhancement of users satisfaction.

b.   POS Tagging

Part-of-speech tagging (POS tagging or POS tagging or POST), also called grammatical tagging or word-category disambiguation. In this process of marking up a word in a text (corpus) as corresponding to a particular part of speech, based on both its definition and its context.

c.   Negation Analysis

In this module, purpose of negation analysis is that analysis of the word based on negative word such as 'not' or 'non' in the reviews. Negations are those words which affect the sentiment orientation of other words in a sentence. Examples of negation words include not, no, never, cannot, shouldn't, wouldn't, etc.

d. Conjunction Analysis

Conjunctions are those words which link clauses with each other in a sentence. Examples of conjunctions include: and, or, but, whereas, etc. Conjunction analysis is the important part of our technique because it is very handy to identify the scope of negation in compound sentence.

e. Tokenization

In the tokenization module, tokenization process is based on sentiment analysis. This analysis based on adjective of the words with negative and positive view.

4. MATHEMATICAL MODEL

Input: Enter opinion feedback (T,S,R,C,D,F)

T: Text input

S: Symbol

R: Remove unwanted symbols

C: Classification

D: Store data base

F: Final Result

Output: Stored user opinion feedback to the Database.

Input: Function Pre-processing (id, request, feedback).

ID: unique id for each feedback.

Request: User request to the server.

Feedback: user opinion feedback

Output: User Feedback will preprocess.

Input: Function Classification (id, feedback)

ID: unique id for each feedback.

Output: Classification of final result will display.

Success Condition: Our system will success when we get the final opinion feedback result.

Failure Conditions: Our system fails when feedback output does not give

6. RESULTS AND ANALYSIS

The proposed system results and graphical analysis is as follows:

(a) User Registration   (b) Admin login

(c)Graphical Visualization

6. APPLICATIONS

Sentiment Analysis has many applications in various fields. The application from a user's standpoint is the applications related to review websites. Application of Sentiment Analysis is important in the automatic summarization of user reviews. Automatic summarization is nothing but the creation of summary of the entire review using an automated program.

7. CONCLUSION

Sentiment analysis is essential for anyone who is going to make a decision. Sentiment analysis is helpful in different field for calculating, identifying and expressing sentiment. It is helpful for everyone when they want to buy a product and they can decide which product is best. Sentiment analysis is very important for Enterprises and helps them to know what customers think about their products. Therefore, companies can take decisions about their products based on customer's feedback. Thus, companies can modify their products features and introduce new products according to customer's opinion in a better and a faster way. Thus, as per the survey done for above techniques, existing techniques are not that efficient as their disadvantages are not being overcome in any existing system. Sentiment analysis is essential for anyone who is going to make a decision. Sentiment analysis is helpful in different field for calculating, identifying and expressing sentiment. It is helpful for everyone when they want to buy a product and they can decide which product is best. Sentiment analysis is very important for Enterprises and helps them to know what customers think about their products. Therefore, companies can take decisions about their products based on customer's feedback. Thus, companies can modify their products features and introduce new products according to customer's opinion in a better and a faster way.

REFERENCES

1. Chenghua Lin, Yulan He, Richard Everson, Member, IEEE, and Stefan Ruger, June 2012,"Weakly Supervised Joint Sentiment-Topic Detection from Text", IEEE transactions on knowledge and data engineering, VOL. 24, NO. 6.

2. Liu,B. "Sentiment Analysis and Opinion Mining," Synthesis Lectures on Human Language Technologies, vol. 5, no. 1, pp. 1-167, May 2012.

3. Rui Xia, FengXu, ChengqingZong, QianmuLi, Yong Qi, and Tao Li, August 2015," Dual Sentiment Analysis: Considering Two Sides of One Review", IEEE transactions on knowledge and data engineering, Vol.27, AnNo.8.

4. Shivaprasad T K and Jyothi Shetty, "Sentiment Analysis of Product Reviews: A Review," International Conference on Inventive Communication and Computational Technologies,(ICICCT 2017).

5. Abbasi, S. France, Z. Zhang, and H. Chen, "Selecting attributes for sentiment classification using feature relation networks," IEEE Transactionson Knowledge and Data Engineering (TKDE), vol. 23, no. 3, pp.447-462, 2011.