



Opinion Mining of Customer Review for Amazon Product

Heenabahen D. Chothani

M.Tech Computer Engineering, Parul Institute of
Engineering & Technology, Vadodara, India

Prof. Sumitra Menaria

Computer Science & Engineering, Parul Institute of
Engineering & Technology, Vadodara, India

ABSTRACT

Now-a-days online shopping will become increasingly important. Merchant sells the product more and more on the internet and many users are using the internet to buy a product and express their ideas or opinions about that product. Many users are using online shopping and before or after shopping they used to read or post review respectively. There are list of review about those products, it is difficult to read all those reviews. That's why there is need of customer to summarize positive & negative opinion so that customer can buy a product easily. There are many online shopping company, I would like to go with Amazon.com for analysis of customer review. Amazon is the largest internet based retailer founded in 1994. Thus my goal is to find most popular and interesting product for customer among a huge amount of product and shows that new products are not always more favorable than old products.

Keywords: *Opinion mining, sentimental analysis, customer review, product aspects*

I. INTRODUCTION

Internet is rapidly growing in past few years. Traditionally people was went to the market and then buy the products. But now people can easily buy small to large product using internet without going market and wasting time. Before buying anything people are used to read review about that product. Everyone is post review regarding that product so it is difficult to read all the reviews and not easy to select that product. Some reviews are positive and some reviews are negative so it creates confusion. So that when people wants to search any product at that time according to review and rating of the product it shows

you product in sequential manner. As customer feedback influences other customer decisions about buying the product, these feedbacks have become an important source of information for businesses when developing marketing and segmenting the customer. Similarly, manufacturers want to read the reviews to identify what elements of a product affect sales most and what are the features the customer likes or dislikes so that the manufacture can target on those areas.

However, it is impractical for customer to manually identify the important and negative aspects of products from numerous reviews. Therefore, an approach to automatically identify the important aspects is highly demanded. Reviews are generally allotted a very high rating or extremely low rating. In such a situation the numerical rating or star rating is not sufficient to highlight the inherent meaning of the review. Some users would like to know the specific features which he or she wants to have in the product before buying the actual product. Generally the classical sentiment analysis mapping the customer reviews or opinion into binary classes – positive or negative, but it fails to identify the product features liked or disliked by the customers or even if there are, they are not provide not explicitly ranking the features both for positive and negative features. There are many online shopping company, I would like to go with Amazon.com for analysis of customer review. Amazon is the largest internet based retailer founded in 1994. Thus my goal is to show the overall features of the product with the help of customer review. So that there is no need of reading all comments at the time to making decision to purchase any product.

II. OBJECTIVES OF THE STUDY

“What other people think” has always been an important piece of content for most of us during the decision-making process. Before long World Wide Web became widespread, many of us asked our friends to recommend an auto mechanic or to explain who they were planning to vote for in elections or consulted Consumer Reports to decide what dishwasher to buy. But the Internet and the web have now made it possible to find out about the opinions and experiences of those in the vast pool of people that are neither our personal acquaintances nor well-known professional critics that is, people we have never heard of. And conversely, more and more people are making their opinions available to strangers via the internet. The recent availability of huge amounts of what is called "content generates by users" on the web, like online shopping. People are used to make review after shopping. Some users would like to know the specific features which he or she wants to have in the product before buying the actual product. For example a user might want to buy a camera having a night vision mode because the majority of the photography is done in the night, and therefore he will try to find a camera having this feature as a top most feature. Generally the classical sentiment analysis mapping the customer reviews or opinion into binary classes – positive or negative, but it fails to identify the product features liked or disliked by the customers or even if there are, they are not provide not explicitly ranking the features both for positive and negative features. Objective of this work

is to give better classification technique than existing work with naïve Bayes classification algorithm.

III. LITERATURE SURVEY

Recently, there has been a wide range of research done on customer reviews. The ongoing research work related to the Opinion mining and Sentiment Analysis are given in this section. It automatically extracts the reviews from the website [1]. It also uses algorithm such as Naive Bayes classifier, Logistic Regression to classify the review as positive and negative review. In [1], the process of opinion summarization has three main steps, such as “Opinion Retrieval, Opinion Classification and Opinion Summarization.” User comments are retrieved from review websites. These comments contain subjective information and they are classified as positive or negative review. Depending upon the frequency of occurrences of features opinion summary is created.

In [2], focuses on review mining and sentiment analysis on Amazon website. Users of the online shopping site Amazon are encouraged to post reviews of the products that they purchase. Amazon employs a 1-to-5 scale for all products, regardless of their category, and it becomes challenging to determine the advantages and disadvantages to different parts of a product. In [3], aims to provide summarized positive and negative features of products, laws or policies by mining reviews, discussions, forums etc.

Table 1 - opinion mining at different levels^[8]

<i>Classification of Opinion mining at different levels</i>	<i>Assumptions made at different levels</i>	<i>Tasks associated with different levels</i>
Opinion Mining at Sentence level	<ol style="list-style-type: none"> 1. A sentence contains only one opinion posted by single opinion holder, there could be multiple opinions in compound and complex sentences. 2. Secondly the sentence boundary is defined in the given document. 	<p>Task 1: identifying the given sentence as subjective or opinionated</p> <p>Task 2: Opinion classification of the given sentence.</p>
Opinion Mining at Document level.	<ol style="list-style-type: none"> 1. Each document focuses on a single object and contains opinion posted by a single opinion holder. 2. Not applicable for blog and forum post as there could be multiple opinions on multiple objects in such sources. 	<p>Task 1: Opinion classification of reviewsClasses: positive, negative, and neutral</p>

Opinion Mining at Feature level.	<ol style="list-style-type: none"> 1. The data source focuses on features of a single object posted by single opinion holder. 2. Not applicable for blog and forum post as there could be multiple opinions on multiple objects in such sources. 	<p>Task 1: Identify and extract object features that have been commented on by an opinion holder.</p> <p>Task 2: Determine whether the opinions on the features are positive, negative or neutral.</p>
----------------------------------	--	--

IV. METHODOLOGY

The ongoing research work related to the Opinion mining and Sentiment Analysis are given in this section. It automatically extracts the reviews from the website. It also uses algorithm such as Naive Bayes classifier, Logistic Regression to classify the review as positive and negative review. In [1], the process of opinion summarization has three main steps, such as “Opinion Retrieval, Opinion Classification and Opinion Summarization.” User comments are retrieved from review websites. These comments contain subjective information and they are classified as positive or negative review. Depending upon the frequency of occurrences of features opinion summary is created.

A. Naïve Bayes Text Classification

The Bayesian arrangement is utilized similarly as a probabilistic strategy as Naive Bayes content classification given in research paper [6]. Utilizing suitable samples which reflect nice, terrible or impartial sentiments, same should recognize the middle of them. Basic feeling demonstrating combines a statistically based classifier with a dynamical model. Those Bayes classifier utilizes single expressions also saying pairs concerning illustration Characteristics. It allocates the input under nice or terrible. The unbiased classes, marks +1, -1 what's more 0 individually. This numerical yield drives a basic first-order dynamical system, whose state speaks to the mimicked enthusiastic state of the experiment's representation.

B. Logistic Regression

Logistic regression has a place with the group of classifiers known as the exponential or log-linear classifiers [7]. Like innocent Bayes, it log-linear classifier works by extricating some set of weighted components from the information, taking logs, and joining them linearly (implying that every element is increased by a weight and afterward included).

In fact, logistic regression alludes to a classifier that characterizes a perception into one of two classes, and multinomial logistic regression is utilized when arranging into more than two classes.

In [3], this paper concentrates on mining reviews from the websites like amazon.com, which allows user to freely write the view. It automatically extracts the reviews from the website. It also uses algorithm such as Naïve Bayes classifier, Logistic Regression to classify the review as positive and negative review. Components of the system are as follows:

a. Text Extraction

After the Login credentials, this module takes the amazon.com URL as the input and extracts all the text from the provide webpage.

b. Source Code Extractor

HTML source code of the webpage is extracted in this module.

c. List of Product

This module will display a list of products from which we have to select a products of our choice to extract review.

d. Display Review List

This module generates the dynamic link and displays all the reviews of the selected product.

e. Stop Word Dictionary

This function contains the stop word list which will be used to eliminate the stop words in the reviews.

f. Algorithm selection

This module allows the user to select any one algorithm that can classify the features of the customer review.

g. Calculate Performance

Once the algorithm is selected the training data is loaded and the performance of the algorithm is measured.

V. PROPOSED WORK

Dataset: Online reviews of Amazon:

Product reviews from Amazon.com containing product data provided by Dredze& Blitzer, 2009 data is in unprocessed form source website [6]. Furthermore the Amazon dataset consists of product reviews for different product types for instance Books, DVDs, Music or Electronics etc. The dataset is given in both forms - unprocessed and pre-processed annotated in negative or positive reviews. However no information was provided in how the data has been pre-processed. Due to the lack of information the pre-processed data cannot be considered and was not utilized as data source. The Amazon reviews are provided as flat files sorted according to product types. For each product type, there exists a folder containing reviews which are again separated into raw reviews in pseudo XML-format and annotated reviews.

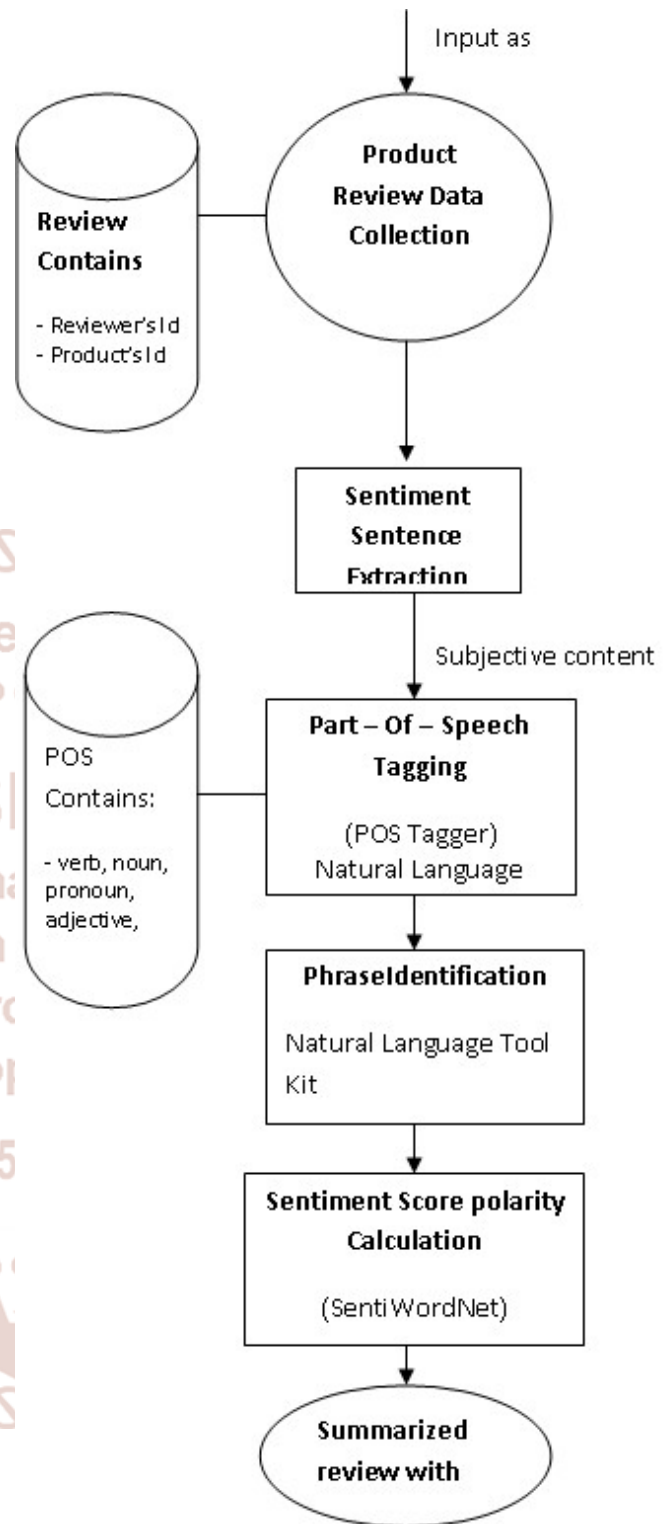


Figure-1: Proposed Work Flowchart

NLTK – the natural language Toolkit is a software tool available for the Python programming language for text processing and mining. Sentiment Analysis or Polarisation is accomplished for instance through tokenisation techniques, POS tagging and calculation of polarity scores.

This research work use NLTK for the programming platform and python as the programming language.

VI. Challenges

1. The challenge is in process of opinion mining that is unstructured and noisy data on website.
2. The human languages are too complex as it changes by person to person. Teaching a machine to analyze various grammatical mistake, misspelling it a very difficult process.
3. Suppose the word is positive in one situation may be negative in another situation. For e.g. Word LONG, suppose if customer says the battery life of Samsung mobile is too long so that would be a positive opinion. But suppose if customer says that Samsung mobile take too long time to start or to charge so it would be a negative opinion.
4. People using social media more and that to for chatting, expressing their views using shortcuts or abbreviations so the use of colloquial words is increased. Uses of abbreviation, synonyms, special symbols is also increase day by day so finding opinion from that is too difficult. For e.g. F9 for fine, thnx for thanks, u for you, b4 for before, b'coz for because, h r u for how are you etc.

VII. CONCLUSION

Opinion Mining or Sentiment analysis refers to extraction of opinion from given text and classify them on the basis of polarity. Opinion mining of customer review is very important to improve service. The customers can make decision rapidly if there is summarized review available from all the review. Automated extract words from a sentence using machine learning method in order to solve different sentiment polarity. Using SentiWordNet which gives collection of words and its polarity, by calculating polarity of each word of amazon data review we can summarized with one opinion whether that particular product contain positive or negative review or which best features are available within the product rather than reading whole long reviews.

REFERENCE

1. Zheng-Jun Zha Member, Jianxing Yu, Meng Wang Member and Tat-Seng Chua, "Product Aspect Ranking and Its Applications," IEEE Transactions on Knowledge and Data Engineering, 2014, Volume: 26, Issue: 5, 1211 – 1224.
2. R. Nithy & D. Maheswari, RVS Coll. of Arts & Sci., Sullur, India, "Sentiment Analysis on Unstructured Review," International Conference on Intelligent Computing Applications, Coimbatore, India, 6 – 7 March 2014.
3. Santhosh Kumar K L & Jayanti Desai, Nitte Meenakshi Institute of Technology Bangalore, India, "Opinion Mining and Sentiment Analysis on Online Customer Review" , International Conference on computational intelligence and computing research, Chennai, India, 15- 16 December 2016.
4. Yuanlin Chen, Yueting Chai, "Analysis of Review Helpfulness Based on Consumer Perspective," IEEE Transactions on Knowledge and Data Engineering, 2015, Volume:20, Issue: 3, 293 – 305.
5. Wenqian Shang, Youli Qu, Houkuan Huang, "A Role-based Customer review Mining System," IEEE International Conference on system , man and cybernatics, Taipei, Taiwan, 8 – 11 October 2006.
6. Neetu, "Hierarchical classification of web content using Naïve Bayes approach", International Journal on Computer Science and Engineering (IJCSE), 2013, Vol-5, Issue: 5, May 2013, 397-975.
7. Nidhi Mishra, "Classification of Opinion Mining Techniques," International Journal of Computer Applications, 2012, Volume -56, Issue-13, 0975 – 888