Abstract

Our day-to-day life has always been influenced by what people think. Opinions of others have always affected our own ideas and decisions. Sentiment Analysis or Opinion Mining is the computational treatment of opinions, sentiments and subjectivity of text. Sentiment analysis or opinion mining is one of the major tasks of NLP (Natural Language Processing). Sentiment analysis has gained considerable attention in recent years. In this paper, we aim to tackle the problem of sentiment polarity categorization, which is one of the fundamental problems of sentiment analysis. A general process for sentiment polarity categorization is proposed with detailed process descriptions. Data used in this study are online product reviews collected from Amazon.com.

Introduction

Due to the rapid growth of electronic commerce, online reviews have replaced the traditional “word-of-mouth” and have been playing a vital role in influencing the consumer’s buying patterns and sales of a product. Reviews act as a trust-building platform for the consumers where by judging the previous buyers’ experience they are able to make informed decisions. From the manufacturer's point of view, helpful online reviews are crucial to mine customer requirements for improving a product or designing a new product. By capturing relevant online reviews, manufacturers can adhere to the customer requirements in the target market. Manufacturers also get an insight to the competitive market and ongoing trends influencing their marketing decisions as well. Retail websites offer different alternatives to the reviewers for writing their reviews. For instance, the user can provide rating in the form of numerical stars (usually ranging from 1 to 5 stars) or open-ended customer authored comments about the product. The presence of online reviews on a website is believed to increase the user credibility, attract consumer visits, augment hit ratio and time spent on the site. The discovery platforms like Zomato and Trivago are booming just on the basis of user reviews provided on restaurants and hotels. Reliable customer reviews build a trust factor among the greenhorn users and help to enlarge the customer base. Both positive and negative reviews help the consumers and the manufacturers. Manufacturers can take negative feedback constructively and can know about the areas that they need to work upon to improve their product or service. In this research, unstructured data is taken into context which will be filtered to remove noisy data and pre-processed to evaluate sentiment of the mobile phone reviews. The proposed work will help future buyers to make better decisions on the basis of analysis of feedback received by a particular smartphone brand. It will also allow manufacturers to meet consumer expectations better on the basis of feedback received.

RELATED WORK

Data Analytics has enabled users to unravel the hidden patterns in data. Big data provides insight on consumer behavior [1] which can be used to make informed decisions. An average consumer is generating both structured and unstructured data which is transforming market decision making. Big Data so generated is defined using three dimensions: Volume, Velocity and Variety. The volume and the relentless rapidity at which data is being generated every day is exceeding the computing capacity of many IT departments. Two more Vs that play an important role in explaining big data are: Veracity and Value. Veracity adds to the
noise and abnormality in data that degrades the quality of data in question. By filtering the irrelevant data, remaining data can be utilized to provide valuable business insights. Big Data has enabled businesses to flourish and improvise on the basis of evidence rather than intuition. It aids in gaining insights on better targeted social influencer marketing, segmentation of customer base, recognition of sales and marketing opportunities, detection of fraud, quantification of risks, better planning and forecasting, understanding consumer behavior, etc. [2]. Sentiment analysis, also known as opinion mining, means identifying the sentiments of the users on the basis of positive, negative and neutral connotations. Sentiment analysis can be classified into three different levels: document level, sentence level and phrase level [3]. A lot of prior research has been done in this field where words and phrases have been classified with prior positive or negative polarity [4]. This prior classification is helpful in many cases but when contextual polarity comes into the picture, the meaning derived from positive or negative polarity can be entirely different. For example, the word ‘amazing’ has a prior positive polarity and the word ‘degrade’ has a prior negative polarity. However, they may be used with negation words like ‘not’ that change the context completely and sometimes phrases containing negation words intensify rather than changing the polarity. For instance, the product delivered was not only good but amazing in terms of looks. This contextual polarity of the phrases was taken into consideration in [5] and ambiguity was removed. [6] used a refined method to establish contextual polarity of phrases by using subjective detection that compressed reviews while still maintaining the intended polarity. Delineated study has been conducted on tweets available on Twitter, movie reviews to build the grounds on sentiment analysis and opinion mining. A sentiment classifier has been built to categorize positive, negative and neutral sentiments not only in English but also for other languages using corpus from Twitter [7]. [8] determined the polarity of smartphone product reviews only on the basis of positive and negative orientation of the review. [9] established a system using support vector machine where sentiment analysis is carried out by taking into consideration sarcasm, grammatical errors and spam detection. Through Statistical and Sentiment Analysis of Consumer Product Reviews (SACP), we have calculated sentence level sentiment orientation of the reviews into ten different sentiments viz. anger, anticipation, disgust, fear, joy, sadness, surprise, trust, positive and negative; providing a much more detailed feedback to the consumer than the previous systems.

The organization of paper is as follows: Section III explicates the dataset used and the approach followed to conduct the analysis. Section IV represents the statistical analysis carried out between different features to determine relationship between them. Section V demonstrates the sentiment of the text of the top three brands. Finally, section VI concludes the proposed work and describes its future scope.

Methodology

Data Pre-processing

For illustrative purpose, I have only considered data with positive sentiment (rating = 4, 5) and negative sentiment (rating = 1, 2, 3) To prepare the data, reviews with ratings greater than 3 will be labelled and encoded as “5” (positive sentiment) and reviews with ratings less than 3 will be labelled and encoded as “1” (negative sentiment Text pre-processing is needed to convert raw reviews into cleaned review. Necessary steps include conversion to lowercase, removal of non-characters, removal of html tags. Depending on machine learning algorithms used, cleaned text data are further transformed to suitable numerical representations

Implementation

Step 1: Preprocess raw reviews to cleaned reviews
Step 2: Create BoW using Count Vectorizer / TfidfVectorizer in sklearn
Step 3: Transform review text to numerical representations (feature vectors)
Step 4: Fit feature vectors to supervised learning algorithm using MultinomialNB and LogisticRegression in sklearn

RESULT & DISCUSSION

Sentiment analysis plays vital role in Businesses and organizations.

- Product and service bench marking.
- Market intelligence. People get the other’s opinion to form decision about product or services.
- Finding opinions while purchasing a new product.
- Finding opinions on political topics.
Sentiment analysis helps to display the product based on the stake holders view in Advertisement (ads).

- Placing ads in the user-generated content.
- Place ads where anyone praises a product.
- Place ads from a close contender if one criticizes a product.

Finally sentiment analysis can be served in the field of Information search & Reclamation. In sentiment analysis determining sentiments appears to be easier, determining objects and their corresponding features is harder. Combining both the task is very tedious and also problem of accuracy.

CONCLUSION
This study described the usefulness of sentiment analysis of in customer reviews in e-commerce domain. In particular, this study demonstrated how Customer review data could be analyzed through a cost-effective application. The study showed that using a relatively straightforward scoring algorithm with a publicly available lexicon will yield a set of sentiment metrics that have reasonable reliability and convergent validity with external customer review portals. As compared to external ranking, the sentiment scores appear to be reasonably effective at capturing real-time information: not only existing customers’ review on their product, but also the public’s buzz toward important short-term events and marketing campaigns. The model has an accuracy of nearly 94%, for prediction of customer reviews.

REFERENCE


