



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 4, April 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118



9940 572 462



6381 907 438



ijirset@gmail.com



www.ijirset.com

Sentiment Analysis of Text Feedback

Harshal Pandharinath Patil, Prajwal Arun Parekh, Tejaswini Rajendra Patil,

Punam Santosh Gangatire

U.G. Student, Department of Information Technology, SSBT College of Engineering and Technology, Bhambhori,

Jalgoan, Maharashtra, India

ABSTRACT: Sentiment Analysis also known as Opinion Mining refers to the use of natural language processing , text analysis to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine. In this project, we aim to perform Sentiment Analysis of product based reviews. Data used in this project are online product reviews collected from “amazon.com”. We expect to do review-level categorization of review data with promising outcomes.

Sentiment analysis or opinion mining is one of the major tasks of NLP (Natural Language Processing). Sentiment analysis has gain much 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.

KEYWORDS: natural language processing , text analysis to systematically identify, extract, quantify, and study affective states and subjective information.

I. INTRODUCTION

Sentiment analysis is a uniquely powerful tool for businesses that are looking to measure attitudes, feelings and emotions regarding their brand. To date, the majority of sentiment analysis projects have been conducted almost exclusively by companies and brands through the use of social media data, survey responses and other hubs of user-generated content. By investigating and analyzing customer sentiments, these brands are able to get an inside look at consumer behaviors and, ultimately, better serve their audiences with the products, services and experiences they offer.

Since customers express their thoughts and feelings more openly than ever before, sentiment analysis is becoming an essential tool to monitor and understand that sentiment .Automatically analyzing customer feedback, such as opinions in survey responses and social media conversations, allows brands to learn what makes customers happy or frustrated, so that they can tailor products and services to meet their customers’ needs.

From a user’s perspective, people are able to post their own content through various social media, such as forums, micro-blogs, or online social networking sites. From a researcher’s perspective, many social media sites release their application programming interfaces (APIs), prompting data collection and analysis by researchers and developers. However, those types of online data have several flaws that potentially hinder the process of sentiment analysis. The first flaw is that since people can freely post their own content, the quality of their opinions cannot be guaranteed. The second flaw is that ground truth of such online data is not always available. A ground truth is more like a tag of a certain opinion, indicating whether the opinion is positive, negative, or neutral. So sentiment analysis using different machine learning algorithms will solve this issue.

The future of sentiment analysis is going to continue to dig deeper, and truly understand, the significance of social media interactions and what they tell us about the consumers behind the screens. This forecast also predicts broader applications for sentiment analysis – brands will continue to leverage this tool, but so will individuals in the public eye, governments, nonprofits, education centers and many other organizations.

1. Scrapping product reviews on various websites featuring various products specifically amazon.com.
2. Analyze and categorize review data. Analyze sentiment on data set from document level (review level).
3. Categorization or classification of opinion sentiment into- Positive Negative.

II. LITERATURE SURVEY

One fundamental problem in sentiment analysis is categorization of sentiment polarity . Given a piece of written text, the problem is to categorize the text into one specific sentiment polarity, positive or negative (or neutral). Based on the scope of the text, there are three levels of sentiment polarity categorization, namely the document level, the sentence level, and the entity and aspect level . The document level concerns whether a document, as a whole, expresses negative or positive sentiment, while the sentence level deals with each sentence's sentiment categorization; The entity and aspect level then targets on what exactly people like or dislike from their opinions.

Relevant literature measure in-textual content statistics in gismo going to apprehend Erik boiy and Marie-Francine moens in their thesis a system mastering approach to sentiment analysis in multilingual internet texts stated that “ Sentiment analysis , to boot mentioned as opinion mining may to be form of facts extraction from matter content of growth analysis and Commercial hobby”. system finding out techniques for sentiment category gain interest because of their practicality to model several capabilities and in doing this, taking footage content their simple ability to changing enter , and therefore the risk to live the certificate a uncertainty by means that of that a class is created.

Sentiment analysis in Czech social media the employment of supervised system mastering ,ivanhabernal ,Tomas Ptacek Associate in Nursinged Josef Steinberger provided an in-depth study of machine learning ways for sentiment analysis of Czech social media . To boot, they provide a proof for the key issue of victimization device learning for sentiment analysis “lies in engineering an adviser set of features ”. Given the case that the majority of the studies in automatic c sentiment analysis of social media has been accomplish English and Chinese , their innovation evolved existing language based entirely sentiment analysis.

Relevant literature review in NPS study is analysis the degree of patron oriented communication and appearance for its relevance the web promoter score and is likewise to provide the power of exchange information processing system.

III. METHDOLOGY

- Reviews

We took the dataset of musical instruments product reviews from Kaggle.

- Data Preparation

Data preparation is the process of cleaning and transforming raw data prior to processing and analysis. It is an important step prior to processing and often involves reformatting data, making corrections to data and the combining of data sets to enrich data.

- Review analysis

Review analysis is the process of transforming unstructured review data to structured data that can be used to guide decision-making

- Sentiment Classification

Sentiment classification is the automated process of identifying opinions in text and labeling them as positive, negative, or neutral, based on the emotions customers express within them.

- Result

The final outcome will be the accuracy score of all three machine learning model.

Dataset Collection

The training of dataset consists of the following steps: Unpacking of data:-The huge dataset of reviews obtained from amazon.com comes in a .csv file format. A small python code has been implemented in order to read the dataset from those files.

Preparing Data

Preparing Data for Sentiment Analysis :-

- 1) The dataset which is loaded contains a lot of coloumns we only need reviewtext and overall rating coloumn for sentiment analysis . All the other coloums are dropped.
- 2) After that we create a binary rating i.e '1' if overall rating is greater than equal to 3 and '0' if less than 3.

Preprocessing Data

This is a vital part of training the dataset. Here we clean the text. We make text lowercase, remove text in square brackets, remove links, remove special characters and remove words containing numbers.

Training Data/ Evaluation

The main chunk of code that does the whole evaluation of sentimental analysis based on the preprocessed data is a part of this.

The following are the steps followed:

- The Accuracy score is calculated and displayed
- Navie Bayes, Logistic Regression, Linear SVM are applied on the dataset for evaluation of sentiments
- Total positive and negative reviews are counted.
- review like sentence is taken as input on the console and if positive the console gives 1 as output and 0 for negative input.

IV. EXPERIMENTAL RESULTS

All the following are the classifier is used for sentiment analysis of text feedback.

• Support Vector Machine

Support Vector Machine (SVM) is a method for the classification of both linear and non linear data. If the data is linearly separable, the SVM searches for the linear optimal separating hyperplane (the linear kernel), which is decision boundary that separates data of one class from another. Mathematically, a separating hyperplane can be written as: $W \cdot X + b = 0$, where W is a weight vector and $W = [w_1, w_2, \dots, w_n]$. X is a training tuple. b is a scalar. In order to optimize the hyperplane, the problem essentially transforms to the minimization of W , which is eventually computed as: $n \sum_{i=1}^n y_i x_i$, where i are numeric parameters, and y_i are labels based on support vectors, X_i . That is: if $y_i = 1$ then $n \sum_{i=1}^n w_i x_i \geq 1$; if $y_i = -1$ then $n \sum_{i=1}^n w_i x_i \leq -1$. If the data is linearly inseparable, the SVM uses nonlinear mapping to transform the data into a higher dimension. It then solve the problem by finding a linear hyperplane.

• Naive Bayesian Classifier

The Naïve Bayesian classifier works as follows: Suppose that there exist a set of training data, D , in which each tuple is represented by an n -dimensional feature vector, $X = [x_1, x_2, \dots, x_n]$, indicating n measurements made on the tuple from n attributes or features. Assume that there are m classes, C_1, C_2, \dots, C_m . Given a tuple X , the classifier will predict that X belongs to C_i if and only if: $P(C_i|X) > P(C_j|X)$, where $i, j \in [1, m]$ and $i \neq j$.

• Logistic Regression

Logistic regression predicates the probability of an outcome that can only have two values (that is a dichotomy). The prediction based on the use of one or more predictors (numerical and categorical).

The following are the result of the project.

In this project we are using three algorithms that is Support Vector Machine, Logistic regression, Naive Bayes.

Classifier	Support Vector Machine	Logestic Regression	Naive Bayes
Accuracy score	0.9570	0.960019	0.9570

V. CONCLUSION

Sentiment analysis deals with the classification of texts based on the sentiments they contain. This article focuses on a typical sentiment analysis model consisting of three core steps, namely data preparation, review analysis and sentiment classification, and describes representative techniques involved in those steps.

Sentiment analysis is an emerging research area in text mining and computational linguistics, and has attracted considerable research attention in the past few years. Future research shall explore sophisticated methods for opinion and product feature extraction, as well as new classification models that can address the ordered labels property in rating inference. Applications that utilize results from sentiment analysis is also expected to emerge in the near future.



The future of sentiment analysis is going to continue to big deeper, far past the surface of thr number of likes , comments and shares , and aim to reach and truly understand , the significance of social media instruction and what they tell us about consumers behind the screens.As a result , sentiment analysis is becoming more important for businesses as the data underlying those interaction grows larger and more complex .

REFERENCES

1. S. ChandraKala1 and C. Sindhu2, "OPINION MINING AND SENTIMENT CLASSIFICATION: A SURVEY,".Vol.
2. Callen Rain,"Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning" Swarthmore College, Department of Computer Science
3. G.Angulakshmi , Dr.R.ManickaChezian ,"An Analysis on Opinion Mining: Techniques and Tools". Vol 3(7), 2014
4. Kim S-M, Hovy E (2004) Determining the sentiment of opinions. In: Proceedings of the 20th international conference on Computational Linguistics, page 1367. Association for Computational Linguistics, Stroudsburg, PA, USA
5. Liu B (2010) Sentiment analysis and subjectivity. In: Handbook of Natural Language Processing, Second Edition. Taylor and Francis Group, Boca
6. Sarvabhotla K, Pingali P, Varma V (2011) Sentiment classification: a lexical similarity based approach for extracting subjectivity in documents. Inf Retrieval 14(3):337–353
7. . Zhou S, Chen Q, Wang X (2013) Active deep learning method for semisupervised sentiment classification. Neurocomputing 120(0):536–546. Image Feature Detection and Description
8. Mukharji A, Liu B, Glance N (2012) Spotting fake reviewer groups in consumer reviews. In: Proceedings of the 21st, International Conference on World Wide Web, WWW '12. ACM, New York, NY, USA. pp 191–200
9. Kristina T (2003) Stanford log-linear part-of-speech tagger. (2014) Scikit-learn. <http://scikit-learn.org/stable>



INNO  SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details