International Journal of Multidisciplinary Engineering in Current Research
ISSN: 2456-4265, Volume 7, Issue 2, February 2022, http://ijmec.com/

SURVEYING SENTIMENTS IN PRODUCT REVIEWS: AN IN-DEPTH ANALYSIS

[1]K. Narayana Rao, [2]Panatula Suresh Babu, [3] Dr. Gowrigari Sasikanth Reddy
[1] Professor, Department of CSE, Rise Krishna Sai Prakasam Group of Institutions, Ongole, [2] Assistant Professor, Department of CSE, Rise Krishna Sai Gandhi Group of Institutions, Ongole, [3] Professor, Department of CSE, Rise Krishna Sai Gandhi Group of Institutions, Ongole.

Abstract:
Sentiment analysis of product reviews has gained substantial attention in recent years due to the exponential growth of e-commerce and the vast volume of user-generated content. This survey paper explores the state-of-the-art methods and trends in product review sentiment analysis, offering insights into the significance, challenges, and potential applications of this field. By examining the existing literature and methodologies, we aim to provide a comprehensive overview for researchers, practitioners, and stakeholders interested in understanding and leveraging sentiment analysis in the context of product reviews.

Keywords: Machine Learning (ML), SVM, Naïve Bayes, Sentiment analysis(SA).

I. Introduction

In the digital age, where consumers have the power to voice their opinions on various products and services through online platforms, understanding customer sentiment and feedback has never been more critical. Product reviews serve as a valuable source of information that can influence consumer purchasing decisions, shape brand perception, and guide product improvement efforts.

Sentiment analysis, a subfield of natural language processing, has emerged as a vital tool for extracting and quantifying sentiments, opinions, and emotions expressed in product reviews. It offers a systematic approach to process the ever-increasing volume of user-generated content, providing valuable insights for businesses, researchers, and consumers alike.

This survey paper is dedicated to exploring the domain of product review sentiment analysis. We will delve into the methods, techniques, and challenges associated with this field, with a focus on its application in real-world scenarios. By the end of this survey, readers should gain a deeper understanding of the current state of sentiment analysis in product reviews and its potential to shape the future of e-commerce and customer engagement.

World Wide Web has become the most popular communication platforms to the public reviews, opinions, comments and sentiments. These sentiments refer to opinions about products, places, books or research papers become daily text reviews. The number of active user bases and the size of their reviews created daily on online websites are massive. There are 2.4 billion active online users, who write and read online around the world. According to 2013 Study 79% of customer's confidence is based on online personal recommendation reviews. As the result, a large number of studies and research have monitored the trending increase of online research resources year by year.

Recently, several websites encourage researchers to express and exchange their views, suggestions and opinions related to scientific papers. Sentiment analysis aims at determining the attitude of a writer with respect to some topics or the overall sentiment polarity of a text, such as positive or negative. Sentiment analysis depends on two issues

International Journal of Multidisciplinary Engineering in Current Research
ISSN: 2456-4265, Volume 7, Issue 2, February 2022, http://ijmec.com/

sentiment polarity and sentiment score. Sentiment polarity is a binary value either positive or negative. On the other hand, sentiment score relies on one of three models. Those models are Bag-of-words model (BOW), part of speech (POS), and semantic relationships. BOW model is the most popular for researchers and based on the representation of terms. The term refers to words in Bag-ofWords model. It neglects language grammar and words ordering. POS tagging is a grammatically tagging model especially verbs, adjectives and adverbs.

Sentiment analysis (also known as opinion mining) refers to the use of natural language processing, text analysis, and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service. The objective of Sentiment Analysis is evaluating the sentiments and opinions of a writer respectively, one topic domain or multi-topic domain. It calculates the aggregate sentiment polarity of online real reviews for one topic based on sentiment classification levels, such as positive or negative. Existing analysis approaches to sentiment reviews can be grouped into four main categories: word level, sentence level, document level, and aspect/ entity level.

2. Literature Survey

There is an approach to use sentiment analysis is with constructing a lexicon with information about which words and phrases are positive and which are negative. For example, SentiWordNet is an overtly obtainable lexical resource in which each WordNet. Synset is ascribed three numerical scores describing how objective, positive, and negative the terms in the synset. This lexicon can either compile manually or be acquired automatically. The annotation of lexical or corpora is usually done by hand, and classifiers are then trained with large sets of features to classify a new batch of words or phrases. There are other approaches to analyze sentiments focus on the mining of sentences or entire documents, rather than to depend on the parity of words. This approach usually works with corpora of text documents. The essential problem with document classification (polarity classification) which is that it has to determine the overall sentiment characteristics of an entire document, while the expressedsentiment can be included in just one sentence or word. In other cases, the sentiment can be expressed implicitly, which makes it even more difficult to detect and classify. However, the context surrounding these 'hidden' sentiments can provide very beneficial information for classifying it. Based on this division of the field of sentiment analysis, we often speak of wordlevel, sentence-level and document-level sentiment classification. On other hand, we find another approach in the mining of sentiment is on the web. Web opinion mining aims at extracting summarize, and track various aspects of subjective information on the Web. This can prove helpful for advertising companies or trend watchers. By a synopsis of Sentiment analysis defection (also called as opinion mining) that refers to the use of natural language processing (NLP), text analysis (TA) and computational linguistics (CL) to identify and extract subjective information in source materials. Sentiment analysis is widely utilized for online reviews and social media for a variety of applications, ranging from marketing to customer service. Importance of sentiment analysis There are millions online users, who write and read online and Internet usage around the world. Online daily sentiments become the most significant issue in making a decision. According to a new survey conducted by Dimensional Research, the survey discusses the percentage of trust online customer reviews as much as personal recommendations. According to 2011 Study 74% of customer's confidence is based on online personal

International Journal of Multidisciplinary Engineering in Current Research
ISSN: 2456-4265, Volume 7, Issue 2, February 2022, http://ijmec.com/

recommendation reviews, 60% in 2012 study, and 57% in 2013 Study. But this percentage increases with respect to 2014 Study: 94% of customer's trust on online sentiment reviews.

3. Problem Statement

Sentiment analysis has been done on a range of topics. For example, there are sentiment analysis studies for movie reviews, product reviews, and news and blogs. Below some general sentiment analysis concepts are discussed.

4. Proposed System

Unsupervised Techniques

In unsupervised technique, classification is done by a function which compares the features of a given text against discriminatory-word lexicons whose polarity are determined prior to their use. For example, starting with positive and negative word lexicons, one can look for them in the text whose sentiment is being sought and register their count. Then if the document has more positive lexicons, it is positive, otherwise it is negative. A slightly different approach is done by Turney (2002) who used a simple unsupervised technique to classify reviews as recommended (thumbs up) or not recommended (thumbs down) based on semantic information of phrases containing an adjective or adverb. He computes the semantic orientation of a phrase by mutual information of the phrase with the word 'excellent' minus the mutual information of the same phrase with the word 'poor'. Out of the individual semantic orientation of phrases, an average semantic orientation of a review is computed. A review is recommended if the average semantic orientation is positive, not recommended otherwise.

Supervised Techniques

The main task here is to build a classifier. The classifier needs training examples which can be labeled manually or obtained from a user-generated user labeled online source. Most used supervised algorithms are Support Vector Machines (SVM), Naive Bayes classifier, Decision Tree and Multi-Layer Perceptron. It has been shown that Supervised Techniques outperform unsupervised techniques in performance (Pang et al, 2002). Supervised techniques can use one or a combination of approaches we saw above. For example, a supervised technique can use relationship-based approach, or language model approach or a combination of them. For supervised techniques, the text to be analyzed must be represented as a feature vector. The features used in the feature vector are one or a combination of the features in below section.

SUPPORT VECTOR MACHINE:

Support Vector Machine (SVM) is primarily a classier method that performs classification tasks by constructing hyperplanes in a multidimensional space that separates cases of different class labels. SVM supports both regression and classification tasks and can handle multiple continuous and categorical variables. The black line that separate the two cloud of class is right down the middle of a channel.The separation is In 2d, a line, in 3D, a plane, in four or more dimensions an a hyperplane. Mathematically, the separation can be found by taking the two critical members, one for each class. This points are called support vectors. These are the critical points (members) that define the channel.The separation is then the perpendicular bisector of the line joining these two support vectors. That's the idea of support vector machine.

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i$$

subject to the constraints:

International Journal of Multidisciplinary Engineering in Current Research
ISSN: 2456-4265, Volume 7, Issue 2, February 2022, http://ijmec.com/

$$y_i\left(w^T\phi(x_i)+b\right)\geq 1-\xi_i \; and \; \xi_i \geq 0, i=1,....,N$$

where C is the capacity constant, w is the vector of coefficients, b is a constant, and represents parameters for handling nonwearable data (inputs). The index i labels the N training cases. Note that y€+-1 represents the class labels and xi represents the independent variables. The kernel is used to transform data from the input (independent) to the feature space. It should be noted that the larger the C, the more the error is penalized. Thus, C should be chosen with care to avoid over fitting.

NAVIE BAYES:

The Naive Bayes algorithm is based on conditional probabilities. It uses Bayes' Theorem, a formula that calculates a probability by counting the frequency of values and combinations of values in the historical data. Bayes' Theorem finds the probability of an event occurring given the probability of another event that has already occurred. Bayes' theorem: Probability of event A given evidence B

Prob (A given B) =Prob (A and B) Prob(A) Prob (A given B) = Prob (A and B) Prob (A) where:

- A (Class) represents the dependent event: A target attribute
- and B(Instance) represents the prior event: A predictors attribute.

Since most of sentiment analysis approaches use or depend on machine learning techniques, the salient features of text or documents are represented as feature vector. The following are the features used in sentiment analysis.

Term presence or term frequency:

In standard Information retrieval and text classification, term frequency is preferred over term presence. However, Pang et al. (2002), in sentiment analysis for movie reviews, show that this is not the case in sentiment analysis. Pang et al. claim that this is one indicator that sentiment analysis is different from standard text classification where term frequency is taken to be a good indicator of a topic. Ironically, another study by Yang et al. (2006) shows that words that appear only once in a given corpus are good indicators of high-precision subjectivity.

POS (Part of speech) Tags:

POS is used to disambiguate sense which in turn is used to guide feature selection (Pang and Lee, 2008). For example, with POS tags, we can identify adjectives and adverbs which are usually used as sentiment indicators (Turney, 2002). But, Turney himself found that adjectives performed worse than the same number of uni-grams selected on the basis of frequency.

5. Result

This section presents the results of the study. The accuracy value shows the percentage of testing data set which was classified correctly by the model. The accuracy of three different machine learning algorithms are shown in below table

	Naive Bayes	SVM	MLP
Product Reviews	74.3 %	94.02 %	98.6 %

Table 1: The accuracy of the machine learning methods on the whole data set

Two algorithms have achieved accuracies over 90% for this product review, although MLP got better results.

6. Conclusion

Sentiment analysis becomes the most important source in decision making. Almost people depend on it to achieve the efficient product. Although, there are hundreds of thousands

of researchers, who write and read online papers daily, the research in this field finds not enough till now. The main goal of this study was to determine which machine learning algorithm of MLP, SVM and Naive Bayes methods performs better in the task of text classification. This was accomplished by using the Amazon beauty products as data set. The classifiers were evaluated by comparing their accuracies in different cases of experiments. In terms of accuracies, MLP tends to do better than SVM, although the differences aren't very large, and the algorithms can reach more than 90% of classification correctly.

References

[1] Zeenia Singla, Sukhchandan Randhawa, and Sushma Jain. Statistical and sentiment analysis of consumer product reviews. In Computing, Communication and Networking Technologies (ICCCNT), 2017 8th International Conference on, pages 1–6. IEEE, 2017.

[2] Priyank Pandey, Manoj Kumar, and Prakhar Srivastava. Classification techniques for big data: A survey. In Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on, pages 3625–3629. IEEE, 2016.

[3] Kasper, W. & Vela, M., "Sentiment analysis for hotel reviews", proceedings of the computational linguistics-applications, Jacharanka Conference, 2011

[4] Vanaja, S. and M. Belwal (2018). Aspect-level sentiment analysis on e-commerce data. 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE.

[5] Ezpeleta, E., I. Velez de Mendizabal, J. M. G. Hidalgo and U. J. L. J. o. t. I. Zurutuza (2020). "Novel email spam detection method using sentiment analysis and personality recognition." 28(1): 83-94.

[6] Chakraborty, K., S. Bhatia, S. Bhattacharyya, J. Platos, R. Bag and A. E. J. A. S. C. Hassanien (2020). "Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media." 97: 106754.