International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 5 ǁ 2022 ǁ PP. 52-56

Machine Learning Based Product Review Polarity Detection for Textual Reviews

Chinmay Ingle, Avinash Sudireddy, Rohit Patil, Prof. Pallavi Ahire
Information Technology Department, Sinhgad Institute of Technology, Lonavala, Pune, Maharashtra, India.

***ABSTRACT**:*
These days there is an expansion in review websites. It has turned out to be considerably more intricate to mine fundamental data from survey sites and take proper choice. Using Natural Language Processing, there is need to identify sentiment of content or document. In this paper Sentiment Analysis is done in view of Rule based mechanism and machine learning approach. both of these strategies are analyzed and discovered that machine learning is most appropriate for Sentiment Analysis in light of the exactness measurement. Sentiment Vader and Senti word net are the Rule based algorithms utilized and LDA analysis on Naive bayes is the machine learning strategy used.

Date of Submission: 08-05-2022 Date of acceptance: 23-05-2022

I. INTRODUCTION

Sentiment analysis or opinion mining is the automated and computational study of people's sentiments, expressions, attitudes and emotions towards a particular target. The target object is any individual, event or topic. The reviews written by people cover these topics. The two terms sentiment analysis or opinion mining can be used interchangeably. However, some researchers stated that both these expressions have slightly different notions . Opinion mining extracts and analyses people's opinion about an entity while sentiment analysis identifies the sentiment expressed in a text then analyses it. Therefore, Sentiment analysis aims to automate the task of finding opinions, identifying the sentiments they express, and then classifying the sentiment polarity. People's opinion plays a crucial role in decision making in various domains. If one wants to buy a particular product, he or she may want to know other's opinion before purchasing the product. In real world, organizations and businesses want to find their consumers feedback about their product or services. In recent years, sentiment analysis applications have spread through many domains from recommendation systems, Ad placements, and trend prediction to healthcare and politics. Recent years witnessed the explosive growth of social media (like blogs, reviews, forums, comments and postings on social networking sites) on the web. Most of the organizations are using these contents to make the decisions The task of mining opinion is formidable due to the need to check the individual web sites. It is very difficult, for a human reader to identify the relevant sites and extract the opinions in them. Therefore, there is need of automated sentiment analysis. Most of the organizations are using their own analysis tools to find the opinions of the consumers. Two approaches are widely used for opinion mining and sentiment analysis: 1) Machine learning based 2) Lexicon based. The machine learning based approach uses various supervised and unsupervised learning algorithms for sentiment classification. While the Lexicon based techniques use a lexicon dictionary with sentiment words related to specific domain for sentiment classification. The dictionary contains the information about the polarity of each word, whether the words are positive or negative. The word in the sentence can be matched with the words in dictionary to determine their polarity. Some researchers combined the machine learning and lexicon based techniques . The various sentiment analysis techniques are categorized as shown in Fig 1.

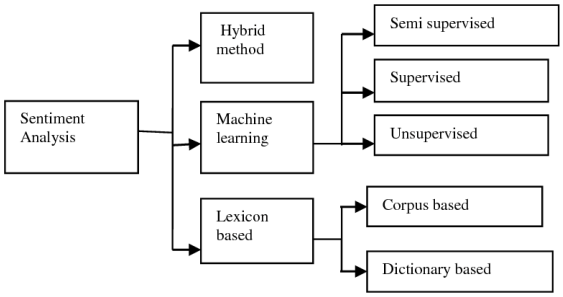


Fig 1: Sentiment Analysis Techniques

II. SENTIMENT ANALYSIS APPROACHES

There are three main classification levels in Sentiment Analysis: document-level, sentence-level, and aspect-level. Document-level sentiment analysis works at document level. It classifies an opinion document as expressing a positive or negative opinion or sentiment. It considers the whole document as a basic information unit. Sentence-level sentiment analysis aims to classify sentiment expressed in each sentence. Before analyzing the polarity of sentiments, there is need to identify whether the sentence is subjective or objective. It determines whether the sentence expresses positive or negative opinions. Both the document level and sentence level sentiment analysis cannot interpret the exact sentiment. Aspect or Feature level sentiment classification does finer-grained analysis . It concerns with identifying and extracting product features from the source data. A substantial work has been done over the past few years in the field of sentiment analysis

Rule-based approach

The main strategy is rules-based and utilizes a dictionary of words named by sentiment to decide the sentiment of a sentence. Sentiment scores regularly should be joined with extra principles to relieve sentences containing negations, sarcasm, or dependent clauses

The NLP techniques which are included in the rules are:

- Stemming, tokenization, part-of-speech tagging
- Lexicons.

Systems involving rule-based approaches are quite simple since the sequential merging of words is not considered. superior processing methods can be utilized and the latest rules can be affixed to support newer forms of expression and vocabularies. But, the addition of new rules can influence previously obtained results and can cause the entire system to become extremely complicated. Frequent fine-tuning and maintenance are required by rule-based systems hence, will also require financing at frequent intervals.

III. MACHINE LEARNING APPROACH

APROACHES	ADVANTAGES	LIMITATIONS
Rule Based Approach	Training data not required High precision Can be a good way to collect data one can start the system with rules let data come by naturally as people use the system.	Lower recall Difficult and tedious to list all the rules
Machine Learning Approach	Dictionary is not required. Exhibit the high precision of classification.	Classifier trained on the textual data in a single field much of the time doesn't work with different fields.
Lexicon Based Approach	Named information and the method of learning isn't needed.	Requires amazing semantic assets which isn't generally accessible.

Machine learning techniques, don't depend on manually crafted rules, but on machine learning procedures. A sentiment analysis task is normally modelled as a classification problem, where the classifier a text data and it returns a class such as positive, negative or neutral

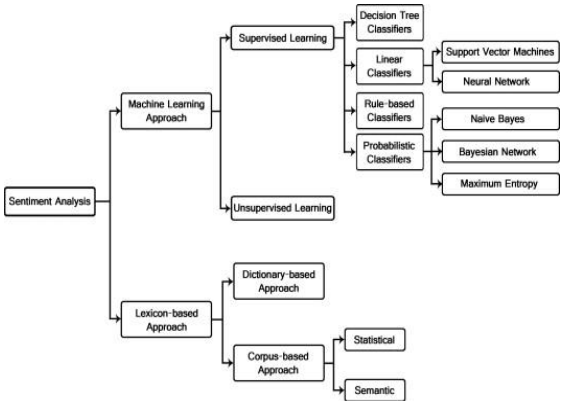
• In the training process, our model learns to compare a particular input data to the respective output data on the basis of test examples utilized for the training process. The feature extractor transforms the textual input into a features vector. Feature tag and vector pairs are then supplied into the algorithm to produce a model.

• In the prediction process depicted in Fig. 2(b), the feature extractor transforms hidden textual inputs into feature vectors. These vectors are then supplied to the model, generating prediction tags for the corresponding vectors.

Lexicon based approach

This method figures the sentiment orientation of the entire text or set of sentences. from semantic direction of lexicons. Semantic direction can be positive, negative, or neutral. The word reference of lexicons can be made manually as well as automatically generated

The methodology for the sentiment categorization task was executed as follows. Firstly, all training weights of text data and the categorized text is calculated. The entire textual data is then stored in a 1-D emotion field. Then the mean weights of the training text data per sentiment category were identified. The categorized text belonged to the category found closer in the 1-D emotion field .



Advantages and Limitations of All the Approaches

In the below Table 1 advantages and limitations between Rule Based approach, Machine Learning

approach and Lexicon-Based approach have been discussed.

Machine Learning Classification Algorithms

Machine Learning Approach vs Lexicon Based Approach., deep learning-based classifiers and transformer-based classifiers .

Conventional Machine learning classifiers

The machine learning models which are commonly used for text classification, which are quite simplistic in their methodology and sort of take a linear approach in classifying the data are considered first for the analysis. The classifiers such as Naïve Bayes, KNN, SVM and Decision Tree are considered.

Naive Bayes (NB)

The Naive Bayes algorithm is one of the intuitive methods among classification algorithms. It is a simple algorithm that makes use of the probability of every feature per category to get respective predictions.This algorithm runs great for the categorization of textual data. It is based on the Bayes Theorem which is used to describe the probability of an event based on its prior knowledge .

Data preprocessing - word Tokenization and NLP

Once the review text gets imported it is considered as each customer feedback which gets extracted in terms of required tokenization and produced a needed relationship by NLP. However, this process perform through NLP has assisted in comprehensively categorized as the controlled program of natural language which may apprehensively in connection among computer and human language from the computer science with deep learning. The large quantity of text has been analyzed and handled with predictive analysis using NLP. This is a part of deep learning technique with some characteristics such as stemming; chunking data and stop words removal get utilized. The beneficial of NLP in creating a sentiment words by segregating the words in term of noun and even the paragraph and sentences are tokenized and chunked in determining the sentences as positive and negative. Thus, the NLP also used for translator in translating one language to needed language. It may generate low noise which may lead to robust data. NLP assist in feeding customer feedback as an input and it get divided into each token using tokenizer. A sequence part of character has combined with organization involving Fpunctuation marks, symbols, special characters, words, etc. that has added in modifying a sentence into various words based on word tokenization. This research has focused with Natural Language Tool Kit (NLTK) is measured and applied with python which get assisted and interpreted to predefine the structure of sentences along its meaning. According to this proposed method, the research need to be modifying the customer feedback represented in the attribute of review text along with text of unstructured to the structured data. At first, the data from part of speech is used in all NLP task for finding noun, verb, adjective and root to each word over the sentence from review text. This proposed chunking NLP algorithm assists in identifying the sentiment words present in the review text such as adverb, noun and adjective that are utilized as a feature which may represent high accuracy

Chunking NLP Algorithm for extracting the required terms

Step: 1 Get extracted in term of required tokenization using Defextract_NN

Step: 2 assign grmr = r “”

Step: 3 NBAR: # Adjectives and Nouns, Noun during terminated words {*}

Step:4 NP:{}# connected with, above, in/of/etc..

Step:5 identifying the opinion words present in the review text such as adverb, noun and adjective {} “”

Step:6 parsing the partial syntactic structure of a sentence Chkr as nltk.RegexParser(grmr) om = set()

Step:7returing over tokenization of specified character cnk as tokrnizerfractory for this chunker

Step:8 for tree in cnk.subtrees(filt = lambda t:t.label() = = 'NP'); om.add(''.join([child in tree.1eaves()for child[0]])) returnom
Step:9 sub= [] forsentenc in data; #extract predefine the structure of sentences along its meaning NN (sentenc)
Step:10concat method in the string class as Sub.append(extract_NN(sentenc)) print (sub) 3.3

Sentiment Analysis

Customer feedbacks are evaluated in this proposed work using SA which has been received from the website. When before charging the money, the customer needs feedback about the company. At the moment to read all the suggestions has not possible which was provided by the customer in the website. However all kind of product analysis or feature analysis present in the companies are available with new information. Therefore, all kind of essential inputs have been provided from the customers are possibly to be missed. Thus the organized review rating frequency has assisted to resolve previous challenges. Then word count has been calculated from the extraction of all tokenized words based on SA. These can be obtained by deep learning. The easiest way to interpret the reviews using an SA along with word count which is to figure out the feedback rating. Hence, the rating can be based on the reviews given by the customer. After the SA output has been received, the consumer should make a quicker and minimized attempt to read the feedback as the decision.The analysis terms are equipped using Document Frequency (DF) or Inverse Document Frequency (IDF) have been used for determining word count are displayed in .

Algorithm for Sentiment Analysis

Step 1: CountVectorizer() converts a collection of text documents to a matrix of token counts
Step 2: assign a shorter name for analyze
Step 3: analyzer = vectorizer.build_analyzer() #which tokenizes the string
Step 4: tokenize the string and continue, if it is not empty If analyzer(s): d = {}
Step 5: Find counts of the vocabularies and transform to array
Step 6: item() transforms the dictionary's (word, index) tuple pairs
Step 7: For k, v in vc.items() D →index:word For index, i in enumerate (w[0]); C →word :count Return C
Step 8: dF1 = dF→ document frequency dF['Rating'] . value_counts(). To_frame()
Step 9: color dF1[Rating] #Rating 4 higher→ positive, Rating 2 lower→negative, Rating 3 → neutral

PERFORMANCE ANALYSIS

The performance of sentiment analysis techniques can be evaluated by using different performance metrics like overall accuracy, precision, recall and F1 score. The values for these metrics can be obtained by using the following confusion matrix.

Table 2. Confusion matrix

Total Samples	Actual Positive	Actual Negative
Classify Positive	TP	FP
Classify Negative	FN	TN

The performance of sentiment classification is evaluated by the Overall Accuracy, which is given by
Overall Accuracy= TP+TN/TP+FP+TN+FN
Another popular evaluation metrics are:

PRECISION is the ratio of the number of relevant records retrieved to the total number of irrelevant and relevant records retrieved. It is usually expressed as a percentage.

Precision= TP/TP+FN

RECALL is the ratio of the number of relevant records retrieved to the total number of relevant records in the database. It is usually expressed as a percentage.

Recall = TP/TP+FP

Applications Of Sentiment Analysis

Brand Sentiment Analysis Management Of Reputation and Social Media Monitoring Reputation Management and Monitoring of Brand are the basic utilization of sentiment analysis over different organizations. Nothing unexpected - knowing on what premise clients notice your organization/stock/administration are for the most part similarly valuable for innovation organizations, marketing agencies, design brands, news, and paper organizations, a few different organizations. As a general rule, utilization of sentiment analysis brings extra adapt ability and the inward delivery of the organization and its items. It empowers organizations to:

Track the comprehension of the organization by clients.Pointing out the specific details of the organization towards the organization.Company should discover the current example and the patterns.Monitor what influencers are recommending about the brand.

IV. Conclusion

The theoretical study concludes that machine learning (Supervised) techniques have performed better in terms of accuracy than lexicon based (Unsupervised) methods in sentiment analysis. However, the performance of machine learning approaches is heavily dependent on the selected features, quality and quantity of training data and the domain of the dataset. The additional training time required by the machine learning approach is also a prominent issue; as lexicon based approach do not require time for training. Therefore, the combined approach of machine learning and lexicon based techniques should be used to avoid the weaknesses of both the techniques and to preserve their strength. The combined approach of machine learning and lexicon based techniques gave the best accuracy of the sentiment classification as compared to the machine learning and lexicon based approaches. It is inheriting high accuracy from machine learning approach and achieving stability from lexicon based approach to enhance the performance of sentiment classification.

ACKNOWLEDGEMENT

I would like to express profound gratitude to my guide Prof Pallavi ahire for her invaluable support, encouragement, supervision and useful suggestions throughout the work.

REFERENCES

[1]. B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," Proceedings of the ACL-02 conference on Empirical methods in natural language processing,vol.10, 2002, pp. 79-86.

[2]. Pang B., Lee, L.,A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts",Proc. Of 42nd annual meeting of the association for Computational linguistics(ACL), 2004, 271-278

[3]. Boiy, E., Hens, P., Deschacht, K., and Moens, M.F. "Automatic sentiment analysis in on-line text" Proceedings Of 11th International Conference on Electronic Publishing(Vienna, Ausrtria) 2007

[4]. Cornard, J.G., and Schilder, F. "Opinion mining in legal blogs" Proc. Of 11th International Conference on Artificial Intelligence and Law (ICAiL'07), ACM,2007, 231-236

[5]. Godbole,N., Srinivasaiah, M.Skiena S. "Large scale sentiment analysis for news and blogs", Proceedings of the International Conference on Weblogs and Social media (ICWSm'07), 2007 [9] B. Pang and L. Lee,"Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval 2(1-2), 2008, pp. 1–135.

[6]. A. Mudinas, D. Zhang, M. Levene, "Combining lexicon and learning based approaches for conceptlevel sentiment analysis", Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining, ACM, New York,NY, USA, Article 5, pp. 1-8, 2012

[7]. Tsytsarau Mikalai, Palpanas Themis. "Survey on mining subjective data on the web". Data Mining Knowledge Discovery 2012; 24:478–514.

[8]. Bing Liu. "Sentiment Analysis and Opinion Mining", Morgan & Claypool Publishers, May 2012.