Journal of Informatics
Electrical and Electronics Engineering, 2024,
Vol. 05, Iss. 02, S. No. 116, pp. 1-8
ISSN (Online): 2582-7006

Sentiment Classification on Mobile Review Using Extraction of Sentiment Conveying Sentences

Mohammad Irsad[1], Ashish Khare[2]

[1]Centre of Computer Education and Training, Institute of Professional Studies, University of Allahabad, Prayagraj, India
[2]Department of Electronics & Communication, JK Institute of Applied Physics and Technology, University of Allahabad, Prayagraj, India
[1]irshad_khan209@yahoo.com, [2]ashishkhare@hotmail.com

How to cite this paper: M. Irsad, A. Khare, "Sentiment Classification on Mobile Review Using Extraction of Sentiment Conveying Sentences," *Journal of Informatics Electrical and Electronics Engineering* (JIEEE), Vol. 05, Iss. 02, S No. 116, pp. 1–8, 2024.

https://doi.org/10.54060/a2zjournals.jieee.116

Received: 08/02/2024
Accepted: 30/07/2024
Online First: 31/08/2024
Published: 31/12/2024

Copyright © 2024 The Author(s). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

In the present time, sentiment analysis has become the most successful technique to identifying people's views, opinions or emotions about any product, service or event. Sentiment analysis became more popular as a result of the widespread use of e-commerce sites and social media platforms such as Twitter, Facebook etc. by individuals who want to express their feelings, views emotions or opinions about any product, service or event. Individuals make efforts in order to express themselves. Sentiment analysis is extremely helpful for companies that are selling a product to determine how their product was perceived by consumers. Therefore, it has become essential to generate fast, reliable and efficient techniques for mining user reviews. In this paper, we have proposed an approach to extract sentiment conveying sentences from the review and used three machine learning classifiers: Random Forest, Multinomial Naïve Bayes and Random Forest. The experimental results show that the machine learning classifiers achieve higher accuracy and Random Forest achieve highest accuracy.

Keywords

Sentiment Analysis, Machine Learning, Sentiment Conveying Sentence Extraction, Sentiment Classification

1. Introduction

The act of expressing thoughts holds important significance not just for individuals but also for automated systems seeking to comprehend opinions. Sentiment Analysis (SA) or Opinion Mining (OM) is the mechanism through which the sentiment or viewpoint of a speaker is ascertained. In the Web 2.0 era, platforms like Facebook (Meta), Twitter (X), e-commerce platforms etc serve as channels for people to articulate their emotions, encompassing happiness, anger, disbelief, sadness, and more. SA/OM involves automated systems that compile and evaluate sentiments and attitudes directed towards various entities, including organizations, products, events, and their specific attributes. With the advancement in web technology, the growth in e-commerce has also increased very rapidly. E-commerce and also social media sites provide a platform to consumers to share their valuable insights in the form of review, comment or feedback which contains important opinions about services and products. As increasing numbers of people utilize social media and other such platforms, study of online information may be used to reflect on changes in people's views, behavior, and psychology (Alamoodi et al. 2021) [1]. Therefore, it has become an important need to develop efficient methods to extract valuable information. This information is crucially useful for consumers as well as businesses. Sentiment classification is one technique for distinguishing between positive and negative reviews. Traditional sentiment analysis approaches, such as BOW (Bag-of-Words) and TFIDF (Term Frequency - Inverse Document Frequency), directly retrieve sentiment from linguistic or semantic features. These methods are used to convert words into non-distributed vectors. To achieve the goal of sentiment analysis, the contextual information, the order and the semantic relationship of the words are not considered by these methods. Therefore, we focus in this research on identifying the word dependency in the text by using dependency parser and extracted the sentences that convey sentiments, followed by supervised approaches for sentiment classification.

The following are the major contributions of this paper:

- Demonstrate the capability of dependency parser to extract sentences expressing sentiments.
- Demonstrate better accuracy of BOW when used with the proposed approach.
- Provides higher accuracy to Random Forest, Multinomial Naïve Bayes Decision Tree classifiers.

The organization of the paper is as follows. Section 2 presents the related work. In section 3, the data set utilized, and the methodology used is described. Result and discussion of the current work is mentioned in section 4. The conclusion of the research work is mentioned in the last section 5.

2. Related Work

In their study, Pang and Lee (2008) [14] presented a concise survey describing different techniques employed in sentiment analysis. The exposition highlighted the significance of sentiment analysis in decision-making processes, elucidated the research challenges inherent in sentiment analysis, and described its applications in different domains such as business intelligence. Moreover, the distinction between sentiment analysis and traditional text mining methodologies was explicated. The survey provided a succinct overview of both supervised and unsupervised classification methods, including subjectivity detection and discrete techniques such as term presence, term frequency, and parts-of-speech for feature extraction in sentiment analysis. Additionally, the treatment of negation handling, the identification of product features from reviews, and the subsequent extraction of opinions regarding these features were discussed. Furthermore, the survey addressed opinion summarization and provided a compendium of datasets commonly utilized by researchers in the field of sentiment analysis.

In their 2018 paper, Amrani et al. [15] presented a hybrid approach to sentiment categorization that combines Random Forest and Support Vector Machine (SVM), two popular machine learning classifiers. This hybrid approach is motivated by the distinct advantages associated with each classifier, and the goal is to synergistically leverage their strengths. Their work employs product review data extracted from Amazon for analysis. The study involves a comparative assessment of the proposed hybrid method against the individual implementations of Random Forest and SVM. The experimental results show that the hybrid approach outperforms the two standalone classifier Random Forest and SVM when evaluated in the context of sentiment classification using Amazon product review data.

A.M. Salem and Ashraf (2020) [13] proposed a model for sentiment classification. They employed five widely recognized machine learning algorithms: Maximum Entropy, K-Nearest Neighbor, Decision Tree, Support Vector Machine (SVM), and Naïve Bayes. The Amazon Mobile Review dataset was used for their experimental evaluation. The experimental result demonstrates that the performance of Naïve Bayes and Maximum Entropy is better than other classifiers.

The n-grams and Bag of Words (BOW) models don't really care about the order of words in a document [4]. BOW is mainly interested in whether a specific word appears or not [6]. Zahoor and Rohilla [2] took a lexicon-based approach to figure out if the sentiment in a piece of text was neutral, negative, or positive, and to calculate overall sentiment scores. On the other hand, Oyebode and Orji [3] used an averaging method to determine the sentiment for each word in a text. Giatsoglou and colleagues [6] tried out different techniques, including lexicon-based methods, word embeddings, and hybrid approaches, to classify online reviews written in both Greek and English. They found that the hybrid method was more accurate, especially when dealing with multilingual content.

Machine learning can be used to build classifiers that identify sentiment in text by converting the text into feature vectors. The process usually involves a few important steps: gathering and cleaning the data, picking out the relevant features, training the classifier on this data, and then seeing how well it performs [5]. Classifiers like Naive Bayes, Support Vector Machines, Logistic Regression, and Random Forest are often used to sort text into specific categories. Many researchers turn to machine learning because it's one of the go-to methods for classifying text.

Classifiers like Naive Bayes, Support Vector Machines, Logistic Regression, and Random Forest are often used to sort text into specific categories. Researchers commonly turn to machine learning algorithms because they're a popular and effective way to classify reviews [7,8]. In this paper, we take a look at and compare three machine learning algorithms: Random Forest, Naive Bayes, and Decision Tree [9,10].

3. Dataset and Methodology

3.1. Dataset Description

The dataset includes 67987 Amazon mobile phone reviews for several international brands. Columns in the dataset include the following attributes.

Asin: For each mobile, there is a unique ASIN (Amazon Standard Identification Number). This column contains the ASIN of all mobiles.

Name: The names of the reviewers are in the name column.

Rating: Consumers rating for mobile is presented by this column. The rating is numeric value and ranges from 1 to 5.

Date: This column contains the date of the review posted on Amazon website.

Verified: This column indicates if a customer is valid or not.

Title: This column is the title of the review.

Body: The detailed review posted by consumer is stored under body column. This column represents the review contents of the reviewer.

Helpfulness: Helpfulness feedback of the review.

3.2. Attribute Selection and Preprocessing

The sentiment classification task focuses on the dataset's "rating" and "body" columns. The "body" column contains customer reviews, while the "rating" column holds the corresponding rating value provided by customers for each review. These ratings range from 1 to 5, where a rating of 1 indicates a very negative opinion of the product, and a rating of 5 indicates a very positive one. The sentiment of each review is inferred from its rating and is then stored in a newly created "Sentiments" column within the dataset. The sentiment can be positive or negative and is calculated as follows:

- Positive: If the rating value is more than 3, it is represented by 1 in the dataset's 'Sentiments' column.
- Negative: If the rating value is 3, it is represented by 0 in the dataset's 'Sentiments' column.

We constructed a tagged dataset for sentiment analysis as a result of this work. The spaCy library is used for data preprocessing, which includes stemming, stop word removal, converting to lower case, punctuation removal, and white space removal.

3.3. Extraction of Sentiment Conveying Sentences

Traditional feature extraction methods, such as TF-IDF and Bag of Words (BoW), have several limitations. These include generating sparse feature vectors and potentially including uninformative features in the text. Additionally, they do not consider the word order within the text.

To solve the aforementioned concerns, the sentences which convey the sentiments are extracted from the review. To perform this task, dependency parser is utilized. Following steps are used to perform this task-

a. **Input:** Raw text reviews.
b. **Tokenization:** Split text into individual words.
c. **POS Tagging:** Label each word's grammatical category.
d. **Identify Sentiment-Bearing POS Tags:** Recognize tags like nouns, adjectives, and adverbs.
e. **Extract Sentiment-Bearing Sentences:** Find sentences with sentiment words.
f. **Output:** Store or return sentiment-bearing sentences from the reviews

3.4. Sentiment Classification

Each review gets a rating from 1 to 5. We label reviews as positive if they score 3 or higher and negative if they score below 3. The dataset is split into 70% for training and 30% for testing. We tokenize the reviews to build a vocabulary, which we use with both the Bag of Words (BoW) and TF-IDF models to turn the text into vectors. These vectors are then fed into the classification algorithms for analysis.

Reviews often have a mix of positive and negative sentences, so we look at all of them to determine the overall sentiment. We use VADER (Valence Aware Dictionary and Sentiment Reasoner) to calculate a sentiment score for each sentence. To figure out the final sentiment of the review, we subtract the total score of negative sentences from the total score of positive ones. If the final score is positive, we label the review as positive; if it's negative, we label it as negative. In our study, we explored three popular machine learning algorithms: Random Forest, Naive Bayes, and Decision Tree.

3.5. Evolution Metrics

Accuracy is a critical factor for a machine learning classifier because the performance of an algorithm in sentiment analysis is characterized in terms of accuracy measurement. Performance evaluation is extremely significant in accuracy measurement [11]. A confusion matrix helps us assess how well an algorithm is performing. In Table 1, you'll see how this matrix works for a problem with two classes. The first column lists the actual class labels, while the first row shows the predicted labels. TP (True Positive) and TN (True Negative) tell us how many positives and negatives were correctly identified. On the flip side, FN (False

Negative) and FP (False Positive) show the number of times positives and negatives were misclassified [12].

Table 1. A Confusion Matrix for a Two-Class Classification [12]

Recognized Actually Class	Predicted as Positive Class	Predicted as Negative Class
Actually Positive class	True Positive(TP)	False Negative(FN)
Actually Negative class	False Positive(FP)	True Negative(TN)

The performance metrics can be defined by using confusion matrix. According to Table 1's confusion matrix, accuracy can be determined as the ratio of accurately expected observations to all observations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision is the ratio of correctly classified positive observations with total positive predicted observations.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall is described as the ratio of the true predicted observations to all of the actual class observations.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

The weighted average of precision and recall is provided by F1-score, which functions on precision and recall.

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{4}$$

4. Experimental Result and Discussion

The experiment has been performed in two steps: in the first step, we performed the extraction of sentiment conveying sentences from the review, calculated the final sentiment value of the review and classified the review in suitable class. This Process is repeated for each review of dataset. In the second step, we trained the models and performed classification using three different machine learning algorithms. The experimental results indicate that the performance of the classifiers is enhanced after applying sentiment conveying sentence extraction approach. Table 2 shows classification performance of machine learning classifiers with BoW. The result shows that Random Forest classifier performs better than other classifiers.

Table 2. Classification result of machine learning classifiers using BoW

Classifier	Class	Precision	Recall	F1-Score	Accuracy (%)
Random Forest (RF)	Negative	0.81	0.69	0.75	88.44
	Positive	0.90	0.95	0.93	
Multinomial Naïve Bayes	Negative	0.75	0.77	0.76	88.17
	Positive	0.92	0.92	0.92	
Decision Tree	Negative	0.65	0.67	0.66	82.99
	Positive	0.89	0.88	0.89	

Table 3 shows the classification performance of classifiers with TFIDF. The result shows that the performance of the Random

Forest and Decision Tree is slightly increased, but the performance of Naïve Bayes classifier is decreased.

Table 3. Classification Result of Machine Learning Classifiers Using TFIDF

Classifier	Class	Precision	Recall	F1-Score	Accuracy (%)
Random Forest (RF)	Negative	0.83	0.69	0.75	88.99
	Positive	0.90	0.96	0.93	
Multinomial Naïve Bayes	Negative	0.84	0.61	0.71	87.52
	Positive	0.88	0.96	0.92	
Decision Tree	Negative	0.66	0.65	0.66	83.19
	Positive	0.89	0.89	0.89	

Table 4 shows the performance of the classifiers after extracting sentiment conveying sentences from the review and performing the classification. The performances of classifiers are much better when compared to applying the classifiers with BoW and TFIDF models. The enhancement in the performance is because the presence of only sentiment conveying sentences in the review and other sentences are removed. It also reduces vector space. Random Forest classifier achieves the highest accuracy of 95.81%.

Table 4. Classification Result of Machine Learning Classifiers Using Sentiment Conveying Sentences Extraction Approach

Classifier	Class	Precision	Recall	F1-Score	Accuracy (%)
Random Forest (RF)	Negative	0.87	0.87	0.87	95.81
	Positive	0.98	0.97	0.98	
Multinomial Naïve Bayes	Negative	0.86	0.78	0.82	94.49
	Positive	0.96	0.98	0.97	
Decision Tree	Negative	0.78	0.87	0.82	93.95
	Positive	0.97	0.95	0.96	

The accuracy of all three classifiers with BoW, TFIDF and proposed approach is shown in figure.

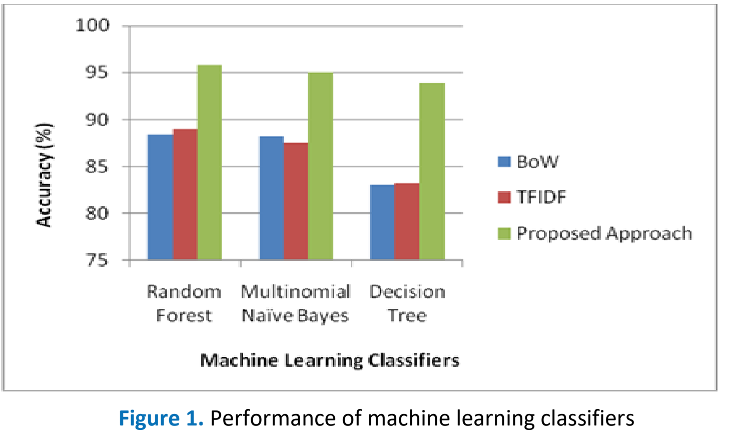

Figure 1. Performance of machine learning classifiers

From the experimental results, it is found that machine learning classifiers perform significantly better with the proposed as compared to the baseline models.

5. Conclusion

The primary goal of this research is to effectively carry out sentiment analysis on an Amazon mobile review dataset making use of the sentiment conveying sentences. We explored several kinds of research problems as well as alternative solutions to problems discovered throughout the sentiment analysis process on reviews. In this paper, we proposed an approach to extract the sentences which convey the sentiments. The proposed approach provides better performance than BoW and TFIDF. The findings from the experiment show that the proposed approach considerably enhances the performance of machine learning classifiers to perform the task of sentiment classification when measured against the baseline machine learning classifiers and other approaches. The future enhancements include expanding the proposed approach for aspect-based sentiment analysis.

References

[1.] A. H. Alamoodi *et al.*, "Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: A systematic review," *Expert Syst. Appl.*, vol. 167, no. 114155, p. 114155, 2021. https://doi.org/10.1016/j.eswa.2020.114155

[2.] S. Zahoor and R. Rohilla, "Twitter sentiment analysis using lexical or rule based approach: A case study," in *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, 2020. https://doi.org/10.1109/ICRITO48877.2020.9197910

[3.] O. Oyebode and R. Orji, "Social media and sentiment analysis: The Nigeria presidential election 2019," in *2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 2019. https://doi.org/10.1109/IEMCON.2019.8936139

[4.] Q. Le, & T. Mikolov, "Distributed representations of sentences and documents", *In Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pp. 1188-1196, 2014.

[5.] O. Y. Adwan, M. Al-Tawil, A. Huneiti, R. Shahin, A. Abu Zayed, and R. Al-Dibsi, "Twitter Sentiment Analysis Approaches: A Survey," *Int. J. Emerg. Technol. Learn.*, vol. 15, no. 15, p. 79, 2020. https://doi.org/10.3991/ijet.v15i15.14467

[6.] M. Giatsoglou, M. G. Vozalis, K. Diamantaras, A. Vakali, G. Sarigiannidis, and K. C. Chatzisavvas, "Sentiment analysis leveraging emotions and word embeddings," *Expert Syst. Appl.*, vol. 69, pp. 214–224, 2017.

[7.] P. Yin, H. Wang, and L. Zheng, "Sentiment classification of Chinese online reviews: analysing and improving supervised machine learning," *Int. J. Web Eng. Technol.*, vol. 7, no. 4, p. 381, 2012.

[8.] R. Xia, T. Wang, X. Hu, S. Li, and C. Zong, "Dual Training and Dual Prediction for Polarity Classification," in *ACL*, 2013, pp. 521–525.

[9.] M. Guia, "Comparison of Naïve Bayes, Support Vector Machine, Decision Trees and Random Forest on Sentiment Analysis"," *KDIR*, 2019 - scitepress.org

[10.] K. Zahoor, N. Z. Bawany, and S. Hamid, "Sentiment analysis and classification of restaurant reviews using machine learning," in *2020 21st International Arab Conference on Information Technology (ACIT)*, pp. 1–6, 2020.

[11.] D. M. E.-D. Mohamed and M. H. N. El-din, "Performance Analysis for Sentiment Techniques Evaluation Perspectives," in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2017*, Cham: Springer International Publishing, 2018, pp. 448–457

[12.] Q. Gu, L. Zhu, and Z. Cai, "Evaluation measures of the classification performance of imbalanced data sets," in *Communications in Computer and Information Science*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 461–471, 2009.

[13.] A. M. Mohammed and Y. A. Ashraf, "Sentiment Analysis of Mobile Phone Products Reviews Using Classification Algorithms," in *2020 International Conference on Promising Electronic Technologies (ICPET)*, 2020. DOI: 10.1109/ICPET51420.2020.00024

[14.] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr*, 2008, doi: 10.1561/1500000011.

[15.] Y. Amrani, M. Lazaar, K. Kadiri, and E. El, "Random Forest and Support Vector Machine based Hybrid Approach to Sentiment Analysis," in *The First International Conference on Intelligent Computing in Data Sciences*, vol. 127, Elsevier, pp. 511–520, 2018.