2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT) | 979-8-3503-9648-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/InCACCT57535.2023.10141741

Detecting Sentiment Polarities with Comparative Analysis of Machine Learning and Deep Learning Algorithms

Thian Lian Ben
Department of Computer Engineering
Marwadi University
Rajkot, India
misterlianben@gmail.com

Ravikumar R N
Department of Computer Engineering
Marwadi University
Rajkot, India
rnravikumar.cse@gmail.com

Poorna Chandra Reddy Alla
Department of Computer Engineering
Marwadi University
Rajkot, India
allapoorna.reddy@marwadieducation.edu.in

Komala G
Department of Computer Engineering
Marwadi University
Rajkot, India
komalag.119556@marwadiuniversity.ac.in

Krishnanand Mishra
Department of Computer Engineering
Marwadi University
Rajkot, India
krishnagec15@gmail.com

***Abstract*— As technology develops, customer reviews are used to assess the quality of products due to the increasing number of selling products online. The extraction of useful minerals from reviews are extremely important for future buyers who are seeking thoughts and sentiments to help their decision-making. Sentiment polarity detection is the process of categorizing the emotions expressed with text, mainly to identify whether the subjectiveness of the writer's attitude toward the product, or service is positive, neutral or negative. To decrease sentiment mistakes on increasingly complex training data, we deploy machine and hybrid learning models that integrate multiple types of deep neural networks in this study. Then, we apply TF-IDF vectorization for extraction of valuable information in reviews. The paper proposes the comparison of performance between machine learning models, deep learning models and the combination of deep learning models to discover the sentiment polarity on online product reviews. We use the dataset collected from an e-commerce website (Amazon), which includes various product reviews. The experimental results display that combination of deep learning models outperform more machine learning algorithms.**

Keywords—sentiment analysis, machine learning, deep neural networks, hybrid learning models, TF-IDF

I. INTRODUCTION

Humans freely express their feelings and emotions on websites. For instance, customers generate opinions of the product on websites such as amazon, yelp, Jingdong, etc. Therefore, online user comments generate a lot of huge data. Due to the comments' fast growth, it is challenging to manually examine them. In the age of big data, it might be useful to quickly grasp network public opinion by extracting the valuable information of comment messages using artificial intelligence technologies. Sentiment analysis is quite beneficial for detecting the sentiment polarity of customer feedback. Sentiment analysis is a text mining approach that encompasses machine learning, natural language processing, text analysis, and other fields of research. Sentiment analysis of customers' reviews typically center on the sentiment of comments, which refer to positive or negative, neutral comments of products. By analysing users' opinions on products, the producers can make the improvement of products' quality. Therefore, sentiment analysis for products' reviews is an important task in businesses, industries and organizations, but it is still a lot of work to capture the users' emotions whether positive or negative, neutral.

Researchers have done sentiment analysis with machine learning and deep learning methods recently. In this paper, we use machine learning methods and the combination of deep learning models. For machine learning methods, support vector machine (SVM), Random Forest classifier and Naive Bayes classifier are applied to get accurate sentiment analysis. These machine learning algorithms are mostly used for text sentiment analysis in the fields of natural language processing. Long short term memory (LSTM), convolutional neural networks (CNN) and bi-directional long short term memory (Bi-LSTM) are applied in deep learning approaches to capture the sequence information and enhance the model's performance [1]. Text classification task is the core of sentiment analysis, and various words contribute in different ways to categorization. Learning a word's low-dimensional, non-sparse word vector representation is an essential step in text categorization tasks. In this study, we use one of vectorization methods, which is known as Term frequency inverse document frequency (TF-IDF) according to the word vector involving the meaning of words and sentiments of words.

The following are the primary contributions of our work:

- We use three machine learning algorithms: SVM, Naive Bayes, and Random Forest classifiers. SVM is used for categorization and finding the best decision boundary between the vectors that associate with a given group or do not associate with it. And Random Forest is intended for feature selection in high dimensional data and classification. Naive Bayes is aimed at solving classification problems, binary classification and multi-class classification as well.

ISBN: 979-8-3503-9648-5/23/$31.00 ©2023 IEEE

- We use CNN, LSTM in deep learning methods and the combination of deep learning models such as CNN + Bi-LSTM and CNN + LSTM. For extracting higher-level sequences of word features, use CNN. Bi-LSTM is proposed to capture the sequence of valuable information and LSTM is utilized to catch long-term dependencies on the word features sequence.
- To strengthen the model and boost its performance, we employ comparison mechanisms.
- We use the amazon polarity dataset, which is collected from hugging face website.

II. RELATED WORK

This [1] paper talks about categorizing the polarity of online products reviews using machine learning techniques. There is a significant research gap: some studies only take into consideration the positive, neutral, and negative sentiment classes, but we can know that the study has considered five classes: extremely positive, positive, neutral, negative, and strongly negative [7]. The dataset was obtained from Amazon using a custom Python-made crawler. A variety of datasets are used in the suggested methodology. Two products' reviews are included in the dataset. 1) Office products 2) Music DVDs. They used the Natural Language Processing Unit (NLTK) for parts of speech (POS) tagging using a particular tag set and Senti-Word Net to score. They compared the performance of six classifiers by utilizing three different feature combinations, including Adverb (RB), Comparative adverbs (RBR) and Adverb + Superlative adverbs (RRS). The classifiers are SVM, Naïve Bayes, and Gradient Boosting classifiers, Sequence to Sequence Model, Decision Tree and Random Forest classifiers. Classifiers with the best accuracy are Random Forest classifiers.

In [2] brand-new sentiment analysis model called SLCABG is proposed for use with online product reviews [2]. The SCABG model is built on the sentiment lexicon and integrates CNN and attention-based Bidirectional Gated Recurrent Unit (BiGRU). To improve the sentiment characteristics in the evaluations, the sentiment lexicon is utilized. The reviews' major sentiment and context characteristics are then retrieved using CNN and the Gated Recurrent Unit (GRU) network and use attention mechanisms for weight. Last, categories the weighted sentiment feature using machine learning algorithms. The paper uses book reviews dataset collected from dangdang.com, a big e-commerce company in China, using website spider and 100000 reviews are included in the dataset, with 50% good ratings and 50% negative reviews. To classify the weighted sentiment features, the paper utilizes machine learning algorithms, including SVM and Naive Bayes Classifiers, and deep learning models such as CNN, combination of CNN and attention mechanism, BiGRU and hybrids between BiGRU and attention weight mechanism and SLCABG. The experimental results suggest that the deep learning model (CNN and BiGRU) classification performance outperforms the machine learning model (NB and SVM).

The purpose of this [3] research is to determine how feature-level ratings of mobile items can be derived from online review and review votes to help manufacturers and new buyers make decisions. A rating system provides a more in-deep perspective of the item than a rating system just at product level. While feature-level evaluations are specific and tell us exactly what is excellent or poor about the product, product-level ratings are too general [4]. They use a data set called "Amazon Reviews: Unlocked Mobile Phones," which PromptCloud retrieved from the Amazon website, to test the suggested strategy. It consists of 4418 number of reviews for mobile phones, product-level ratings, and voting's. The collection of data includes 413 841 reviews in all, along with the votes each one received.

III. PROPOSED WORK

The steps of proposed work are as shown in figure 1) Before extracting the features from the input text, data pre-processing is employed first because the reviews may include emoji, icons and improper words. In machine learning, data preparation is a process of transforming (normalizing and cleaning) the review data into something useful that can be applied to build and train machine learning models. In this stage, there are some common preparation/cleaning steps such as tokenization, lemmatization, removal of punctuation, converting emoji to words, removal of stop words, spelling corrections, POS tagging, and converting lower case. Tokenization is the first step of data pre-processing. It is a process of splitting the sentence into words based on specific delimiter. Then, it will remove punctuations included in the raw data. And eliminating the emoticons would not be a smart idea because they provide some essential information in use cases like sentiment analysis. So converting emoticons to words will help the machine to understand human emotions. Stop words are frequently used term words in a language. Stop words in English include "a," "the," "is," "are," and others. They can be eliminated from the review data because they don't carry valuable information. The next step is spelling correction. The reviewers may type wrong words. Correcting those words will give the meaningful information of the review. Then, POS tagging is a method of labelling the words with its suitable part of speech. It includes nouns, verbs, adjectives, adverbs and their sub-categories. A step of data pre-processing is converting lowercase is used for reducing the duplication of words like 'product' to 'PRODUCT'. The last step is lemmatization. It returns the base word of the language like from 'talking' to 'talk'.

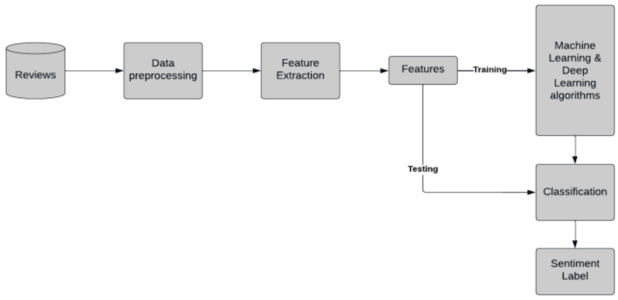

Fig. 1. Proposed work flow

The next process is feature extraction, which uses the TF-IDF algorithm and it will produce the features of text. Plus, the model is trained using machine learning algorithms and deep learning models. Finally, the model has to classify the label of text.

IV. METHODS

In this study, we use some methods for classifying the sentiment of opinion words from customers. The methods used in this study are described here.

A. Machine Learning methods

SVM is a method of classification algorithms which can be used to determine positive or negative sentiments. It is also used for regression problems. After the process of transforming texts into vectors using text vectorization models, the algorithm helps to find the best decision boundary between the vectors that associate with a given group and vectors that do not associate with it. Naive Bayes classifier is used to forecast the probability that a specific record or data point falls into that category. It can perform a straightforward and effective machine learning classification problem. Random Forest is also used in text classification methods that evaluate the likelihood for each class by combining different decision trees.

B. Hybrid Methods

In this study, we use deep learning models and hybrid methods to achieve better performance. Hybrid methods combine some deep learning models [10]. Many techniques can be used to construct a hybrid model for sentiment analysis. Figure 2 presented the process of hybrid deep learning models. We begin with TF-IDF vectorizer to generate the feature vectors and then we go to deep learning models and the combination of deep learning models:

1) TF-IDF -> CNN -> Relu

2) TF-IDF -> BiLSTM -> Relu

3) TF-IDF -> LSTM -> Relu

4) TF-IDF -> CNN -> LSTM -> Relu

5) TF-IDF -> CNN -> BiLSTM -> Relu

TF-IDF stands for Term Frequency-Inverse Document Frequency, it's a numerical statistic that is intended to reflect how important a word is to a document in a corpus [11]. It is often used as a weighting factor in information retrieval and text mining. The TF part of the name represents the term frequency, which is the number of times a term appears in a document, and the IDF part represents the inverse document frequency, which is the logarithmic measure of the frequency of a term in the corpus. The resulting TF-IDF score is used to rank the importance of terms in a document relative to an entire corpus of documents [12]. TF-IDF is used in our experiments for information retrieval. And the next step is deep learning models and hybrid methods.

CNN stands for Convolutional Neural Network, a type of artificial neural network used for image and video recognition, classification, and object detection tasks in computer vision. CNN is one of several varieties of artificial neural networks, which are utilized for diverse applications such as image classification, etc. It is composed of input layer, output layer and several hidden layers and it has numerous parameters [13]. Therefore, it can produce good performance on image recognition. Likewise, it can do well in terms of sentiment analysis. In this paper, we use one convolutional layer.

BiLSTM stands for Bidirectional Long Short-Term Memory, a type of recurrent neural network (RNN) used for sequence modeling tasks such as natural language processing and speech recognition. It processes input sequences in two directions (forward and backward), allowing for the capture of both past and future context for each input element in the sequence [14]. BiLSTM includes bidirectional recurrent neural networks (BiRNN) and LSTM. Bi-LSTM is utilized for gathering sequence data to obtain the text's order and enhance the model.

ReLU stands for Rectified Linear Unit, a type of activation function commonly used in deep learning neural networks. It replaces all negative inputs with zero, allowing the network to learn non-linear relationships in the data.

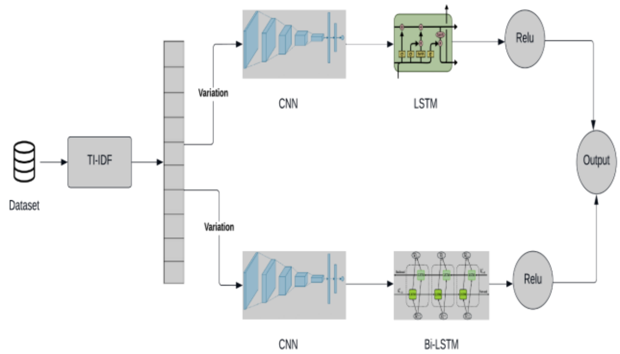

Fig. 2. Process of hybrid deep learning model

V. EXPERIMENT RESULTS

The experimental dataset is amazon reviews dataset, which includes the reviews from Amazon which is selected from Hugging Face. The data cover an 18-year period and contain 36 million reviews as of March 2013, of which 18 millions are positive comments and the remaining are negative comments . The dataset consists of product, customer data, ratings and comments.

The sample five raw data are displayed in Table 1 (where label 1 is negative polarity and 2 is positive polarity).

TABLE I. AMAZON REVIEWS DATASET EXAMPLE

ID	Label	Title	Content
1	2	Great Read	This novel was fantastic, but also realistic in my opinion. That demonstrated to me that human mistake is unavoidable. I liked how this author depicted God's compassionate side rather than his furious side. I couldn't put it down because I enjoyed how it twisted and spun. The Glass Castle was another favourite of mine.
2	2	Barbie as Rapanzel: A creative adventure	I bought this programme for my 5-year-old granddaughter, and she adores it. She occasionally lets me play, but she doesn't always approve of what I come up with. I intend to acquire other software products for her.

3	1	letters from home	I like letters from home because it tells a story if you really listen you can actually hear john Michael montogomery tell his story of soldiers and all that stuff I liked it and I like punk rock and rock so it must be good if I like it
4	1	This is a book for Age 4-8!	I missed that fact before ordering. If you're looking for a children's book this one is well done.

In this experiment, Naive Bayes, Random Forest and SVM are selected as machine learning algorithms. As deep learning models, CNN, LSTM, and BiLSTM and then the combination of CNN - LSTM, and the combination of CNN-BiLSTM are tested as hybrid methods. The results are shown as table 2:

TABLE II. EXPERIMENTAL RESULTS

Algorithms	Accuracy
Naive Bayes	75%
Random Forest	74.5%
BiLSTM	78%
LSTM	80.67%
CNN-BiLSTM	80.13%
Support Vector Machine	82%
CNN-LSTM	82.23%
CNN	**85%**

From the result table, it shows that SVM can perform with the highest accuracy in machine learning models. Naive Bayes classifier is 0.5% better than Random Forest. In the results of deep learning models, CNN is significantly better than other deep learning models. It can be seen that hybrid models can perform more than BiLSTM of deep learning models. There are many feature extraction methods and TF-IDF is used in this study. In comparison to machine learning models, deep learning has a better capacity for learning because it can more clearly capture the data's regional features. The model that performs the learning ability with the highest accuracy is CNN for deep learning and SVM for machine learning models. Furthermore, it can be seen that CNN-LSTM is stronger than CNN-BiLSTM.

VI. FUTURE DIRECTION

Sentiment analysis with emoticons has grown in prominence in recent years since the popularity of emojis in modern media has increased. There is more potential for developing complete emoji libraries, conducting cross-lingual mood analysis, and comparing emoji and text. Algorithms such as rule-based approaches, emotion lexicon-based approaches, emoji embedding-based approaches, and hybrid machine learning approaches that combine two or more distinct algorithms can be investigated further to accomplish these tasks.

VII. CONCLUSION

In this day and age of internet usage growth, it is extremely crucial to study the subjective tendencies of customers' feedback using artificial intelligence technology. Opinion mining, called sentiment analysis, is used to explore the emotional words which are muted in text. This paper has described the comparative study of ML and DL algorithms for sentiment analysis on product reviews. In this study, there are three major steps. The first step is data preprocessing, including removal of stopwords, conversion emojis to words, removing punctuations and POS tagging. This step is one of the most challenging NLP problems because there are no particular statistical criteria. The second step is feature extraction, which uses the TF-IDF method. To compute the weight of the words, the sentiment information is included into the standard TF-IDF method. As a result, the term vector might be more accurately expressed. Lastly, ML algorithms and DL models are used to train the model to determine the probability of posteriority based on previous information and data. According to the experimental results, the comparative study shows that CNN is performing better than CNN-LSTM, SVM CNN-BiLSTM.

REFERENCES

[1] G. Xu, Y. Meng, X. Qiu, Z. Yu and X. Wu, "Sentiment Analysis of Comment Texts Based on BiLSTM," in IEEE Access, vol. 7, pp. 51522-51532, 2019, doi: 10.1109/ACCESS.2019.2909919.

[2] L. Yang, Y. Li, J. Wang and R. S. Sherratt, "Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," in IEEE Access, vol. 8, pp. 23522-23530, 2020, doi: 10.1109/ACCESS.2020.2969854.

[3] Y. Lin, J. Li, L. Yang, K. Xu and H. Lin, "Sentiment Analysis With Comparison Enhanced Deep Neural Network," in IEEE Access, vol. 8, pp. 78378-78384, 2020, doi: 10.1109/ACCESS.2020.2989424.

[4] K. R. Jerripothula, A. Rai, K. Garg and Y. S. Rautela, "Feature-Level Rating System Using Customer Reviews and Review Votes," in IEEE Transactions on Computational Social Systems, vol. 7, no. 5, pp. 1210-1219, Oct. 2020, doi: 10.1109/TCSS.2020.3010807.

[5] W. Meng, Y. Wei, P. Liu, Z. Zhu and H. Yin, "Aspect Based Sentiment Analysis With Feature Enhanced Attention CNN-BiLSTM," in IEEE Access, vol. 7, pp. 167240-167249, 2019, doi: 10.1109/ACCESS.2019.2952888.

[6] B. K. Shah, A. K. Jaiswal, A. Shroff, A. K. Dixit, O. N. Kushwaha and N. K. Shah, "Sentiments Detection for Amazon Product Review," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-6, doi: 10.1109/ICCCI50826.2021.9402414.

[7] S. Kausar, X. Huahu, W. Ahmad, M. Y. Shabir and W. Ahmad, "A Sentiment Polarity Categorization Technique for Online Product Reviews," in IEEE Access, vol. 8, pp. 3594-3605, 2020, doi: 10.1109/ACCESS.2019.2963020.

[8] R. Man and K. Lin, "Sentiment Analysis Algorithm Based on BERT and Convolutional Neural Network," 2021 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 2021, pp. 769-772, doi: 10.1109/IPEC51340.2021.9421110.

[9] A. Poornima and K. S. Priya, "A Comparative Sentiment Analysis Of Sentence Embedding Using Machine Learning Techniques," 2020 6th International Conference on Advanced Computing and

Communication Systems (ICACCS), 2020, pp. 493-496, doi: 10.1109/ICACCS48705.2020.9074312.

[10] W. Meng, Y. Wei, P. Liu, Z. Zhu, and H. Yin, "Aspect Based Sentiment Analysis with Feature Enhanced Attention CNN-BiLSTM," *IEEE Access*, vol. 7, pp. 167240–167249, 2019, doi: 10.1109/ACCESS.2019.2952888.

[11] B. K. Shah, A. K. Jaiswal, A. Shroff, A. K. Dixit, O. N. Kushwaha, and N. K. Shah, "Sentiments Detection for Amazon Product Review," in *2021 International Conference on Computer Communication and Informatics, ICCCI 2021*, Jan. 2021. doi: 10.1109/ICCCI50826.2021.9402414.

[12] R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, "The Impact of Features Extraction on the Sentiment Analysis," *Procedia Comput. Sci.*, vol. 152, pp. 341–348, 2019, doi: https://doi.org/10.1016/j.procs.2019.05.008.

[13] R. Man and K. Lin, "Sentiment analysis algorithm based on bert and convolutional neural network," in *Proceedings of IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers, IPEC 2021*, Apr. 2021, pp. 769–772. doi: 10.1109/IPEC51340.2021.9421110.

[14] W. Yue and L. Li, "Sentiment Analysis using Word2vec-CNN-BiLSTM Classification," pp. 3–7.