Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 47 (2026), No. 5-B, pp. 2003–2010
DOI : 10.47974/JIOS-2291

Optimized multi-class sentiment classification of Flipkart product reviews using deep learning

Sonica Upadhyay *
Yogesh Kumar Gupta †
Department of Computer Science
Banasthali Vidyapith
Newai 304022
Rajasthan
India

Abstract

Due to the exponential growth of textual information on the internet, online monitoring and mining of textual data have become a prominent task for researchers, which requires a deeper understanding of text classification algorithms. Several machines and deep learning algorithms have performed well in natural language processing for textual classification. These classification algorithms extract helpful information from textual resources and automatically classify them into multiple predefined categories based on their content and subject matter. In this research paper, a comparative review sentiments analysis of flip kart products has been done by using different deep learning algorithms like Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized BERT Approach (RoBERTa) on a given dataset in which their efficiency is analysed and compared. and from obtained experimental RoBERTa model scored highest prediction accuracy, with 86.94%.

Subject Classification: *Primary 68T01, Secondary 97C30.*

Keywords: *Deep learning, Traditional models, Text classification, Evaluation metrics, Machine learning.*

1. Introduction

Size of textual data is increasing daily on social media platforms and very difficult for researchers to abstract functional pieces of information from such significant, unstructured data sources. Several textual

* *E-mail:* **s40847819@gmail.com (Corresponding Author)**
† *E-mail:* **gyogesh@banasthali.in**

classification algorithms have been performing well in extracting usable structured information from unstructured data sources and classify textual information into their pre-defined text domain classes. In the last few years, the demand for automatic text classification has also increased due to the generation of many text documents and more structured the data information, the better the analysis and the better the decisions. To achieve this, NLP models are applied for pre-defined tasks or rules, use deep learning classifiers for better content or intent analysis of messages, and finally interpret the structured underlying meaning. This research paper comprises a comparative text classification review analysis of customer sentiment by using some deep learning classifiers based on Flipkart product datasets. The final extracted discriminative feature vector information is fed to the deep learning classifiers (GRU, LSTM, BERT, and RoBERTa) for classifying reviews as positive, negative, and neutral. This text classification approach evaluates all deep learning classifiers' effectiveness and accuracy using F1-scores, precision, accuracy, and recall score metrics [1], [2].

2. Methodology

Sentiment analysis research mainly aims to identify customers' opinions and suggestions regarding comments or reviews about different products or events on social media platforms. This research paper examines customer product reviews' opinions on the Flipkart products dataset and analyses experimental sentiment outcomes.

2.1 *Data Collection & Data Pre-Processing*

Data is extracted from the Kaggle online repository (https://www.kaggle.com/datasets/niraliivaghani/flipkart-product-customer-reviews-dataset/data). It consist 205052 rows and 06 columns with five key input features name, price, rating, summary, and review, and one output feature is sentiment for reviewing customer reviews The main task of the research is to predict which reviews about customer products are positive, negative and neutral. The distribution of key values of output class 'sentiment' has three different sentiment classes, like 'positive', 'negative', and 'neutral' sentiments, which are 166581, 28232, and 10239, respectively are shown in Figure 1.

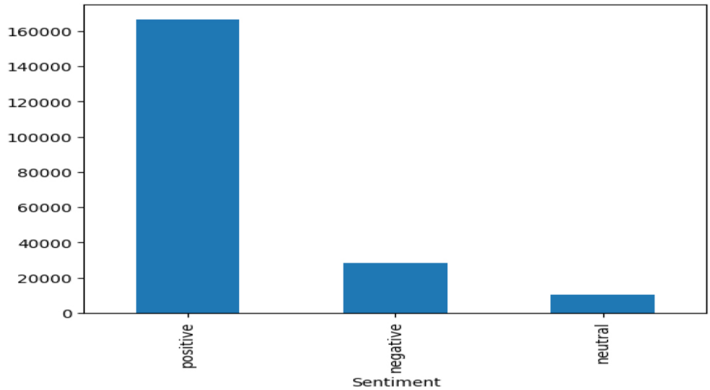

Figure 1

Sentiments Key Values Distribution

Data Pre-processing is the next step for an NLP model after labelling and collecting review sentiments. Removing of null values, stop words, punctuation, and some special characters from the labelled dataset is primary function of data preprocessing [3], [4].

2.2 *Tokenization & Embedding*

Tokenization is primary part of any natural language processing (NLP) problem. It break body's text structure into small units called tokens and these small tokenized units are analysed individually as character and sequences or many more, depending on the type of tokenization shown in Figure 2 [5], [6].

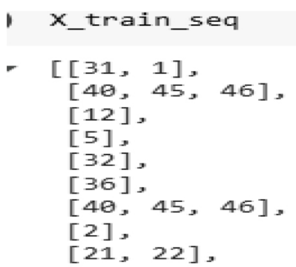

Figure 2

Word Tokenization of Textual Data

Embedding is used to assigns numerical vector values to all individual tokenized words or phrases that help make it easier for the computer to understand and finally the text sequence length is managed by padding given in Figure 3 [7], [8].

```
X_train_pad

array([[31,  1,  0, ...,  0,  0,  0],
       [40, 45, 46, ...,  0,  0,  0],
       [12,  0,  0, ...,  0,  0,  0],
       ...,
       [33,  0,  0, ...,  0,  0,  0],
       [15, 42,  0, ...,  0,  0,  0],
       [54, 14,  0, ...,  0,  0,  0]], dtype=int32)
```

Figure 3

Padding of Textual Data

3. Model Descriptions

The existing Recurrent Neural Network (RNN) model is good for sequence2sequence converter but no good for long-range dependencies but on the other hand transformer models are solved this problem very smartly and used both encoder and decoder to manage long dependencies, in which the encoder is used to read the input text and the decoder to predict tasks [9], [10].

3.1 *BERT (Bidirectional Encoder Representations from Transformers)*

BERT was new language encoder-based transformer representation model that used pre-training and fine-tuning to create various NLP tasks. BERT model helps machines to understand or enable the intricate nuances and contextual dependencies of comprehending language, making human communication richer and more meaningful. The architecture of the BERT model is based on the self-attention mechanism that allows BERT to assign importance to individual words based on their context is shown in Figure 4. The input text value is supplied to the encoder as a sequence of tokens, converted into vectors, and then processed in the neural network. But before the start of processing, some extra metadata bits [CLS] and [SEP] are added to start or end of first input sentence, and process is known as embedding [11].

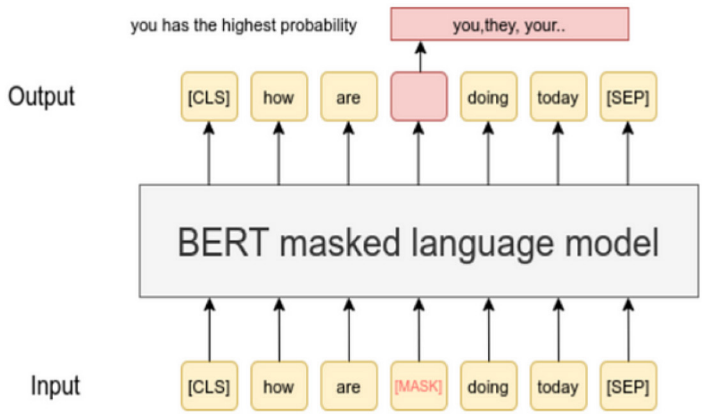

Figure 4

BERT Architecture

3.2 *LSTM (Long Short-Term Memory)*

LSTM is advanced version of Recurrent Neural Networks (RNN). LSTM model also solves sequence-based problems like the RNN model. LSTM model overcomes the problem of short-term memory issues of RNN by retaining information over long distances. LSTM's architecture deals with both long-term (LTM) and short-term (STM), which helps make all calculations simple and effective. It uses the concept of gates. LTM information enters from the forget gate and forgets all useless information and current input, and STM (recently learned necessary information) is combined and applied as the current input. LTM and the combined information of both STM and Event are combined and entered into the Remember gate, which is used to update LTM. Finally, the gate, which uses all three pieces of information from LTM, STM, and Event, works to predict the output of the current event, which works as an updated STM [12].

3.3 *GRU (Gated recurrent unit)*

GRU has some advanced features compared to LSTM, which was introduced in 2014. GRU is similar to LSTM and it also uses gates for managing flow of information. The GRU has advanced features than the LSTM and RNN models. The training time of GRU is faster than that of other models, which makes it computationally less expensive. The GRU model is more effective for solving sequential task problems and is less prone to gradient problems. The architecture of GRU is also straightforward compared to LSTM because it does not require any separate cell state (Ct). The GRU model uses a hidden state and is also faster to train [13].

3.4 *RoBERTa (Robustly Optimized BERT-Pretraining Approach)*

After Google released the BERT model in 2018, proposed a new robustly optimized approach model called RoBERTa, an improved version of BERT. The RoBERTa model was re-implemented from the BERT model by adding new key hyper parameters and embedding. The RoBERTa model does not need to assign token type IDs for dividing the segments like the BERT model and it uses a tokenizer.sep token that easily separates the segments. The architecture of the RoBERTa model is the same as the BERT model. The RoBERTa model was used byte-level BPE as a tokenizer for pre-training and was trained on a large, uncompressed, massive text dataset of 160 GB [14].

4. Experimental Result

The performance and experimental results of various deep learning text classification models were evaluated and implemented using a Python with specification of 2GB RAM, 4 TB of hard disk and an Intel Core i9 processor of computer system. This paper evaluates the performance of various deep learning text classifiers on the Flipkart dataset with prediction accuracy, recall, precision, and F-score. RoBERTa model scored highest accuracy among all models, with value 0.8694, and LSTM model scored the minimum accuracy score, among all four models, with accuracy value 0.3227. Numerical comparative accuracy analysis is given in Table 1.

Table 1

Comparative Accuracy Analysis

S. No	Epochs	Accuracy			
		GRU	LSTM	BERT	ROBERTa
1	Epoch1	0.5718	0.3308	0.855794	0.860352
2	Epoch2	0.6343	0.3319	0.861654	0.869141
3	Epoch3	0.6435	0.3277	0.863281	0.870117
4	Epoch4	0.6352	0.3359	0.862956	0.873372
5	Epoch5	0.6437	0.3227	0.861328	0.869466

5. Conclusion

In recent years, text sentiment analysis has become the most significant human challenge, and all NLP researchers are trying to find an optimized solution for reviewing customers' sentiments. In this research article, an experiment is conducted on the Flipkart product dataset and tries to find out the best model that provides the most accurate or predicted results by using various deep learning text classifiers for analyzing customer sentiment to determine the general opinion (sentiment) of the people about the product. Firstly, suitable or appropriate data is collected from an online data repository, and then data labelling and pre-processing are carried out to improve the quality of the collected data. The exploratory investigation and feature extraction are also performed to extract the discriminative feature vectors that are finally fed into deep learning classifiers for customers' sentiment prediction. The final experimental result showed that the RoBERTa model achieved the highest accuracy score and performed better for customer sentiment prediction. The accuracy score of various deep learning classifiers is evaluated by score matrices like prediction accuracy, recall, precision, and F1-score. In the Flipkart customer sentiment analysis, the RoBERTa model achieved % highest accuracy score of 86.94 % among all deep learning classifiers. Therefore, as a future extension, a novel deep learning optimization technique can be included and experimented with the result on a large dataset, to improve the accuracy of the prediction text classification.

References

[1] A. Rohanroy, K. Rathor, and B. Jannawat, "Reviewing Flipkart product comments using methods based on sentiment analysis," *Int. Res. J. Modernization Eng., Technol. Sci.*, vol. 5, no. 2, pp. 800–806 (2023).

[2] Y. Uttam and D. Sandeep, "Aspect-based sentiment analysis on Flipkart data," *Int. J. Res. Appl. Sci., Eng. Technol.*, vol. 11, no. 5, pp. 5489–5503 (2023).

[3] A. Godia and L. K. Tiwari, "Sentiment analysis and classification of product reviews: A comprehensive study using NLP and machine learning techniques," in *Proc. 10th Int. Conf. Adv. Comput. Commun. Syst.*, pp. 1247–1252 (2024), doi: 10.1109/ICACCS60874.2024.10717296.

[4] K. S. Kumar, B. Bharathi, and K. Susmitha, "Advanced machine learning based aspect level sentiment analysis for Flipkart products," *Int. J. Creative Res. Thoughts*, vol. 11, no. 4, pp. 348–352 (2023).

[5] M. S. Lakshmi, S. P. Kumar, and M. Janardhan, "Machine learning centric product endorsement on Flipkart database," *Int. J. Eng. Adv. Technol. (IJEAT)*, vol. 9, no. 1, pp. 1583–1586 (2019).

[6] V. Shukla and S. Bhoite, "Sentimental analysis of customer reviews by using data analysis," in *Proc. 3rd Natl. Level Students' Res. Conf. Innovative Ideas Inventions Comput. Sci. IT Sustainability,* vol. 8, no. 6, pp. 307–315 (2022).

[7] B. M. Munaf, R. Ansari, and S. O. A. Khan, "Flipkart reviews sentiment analysis using Python," *Int. J. Res. Publication Rev.*, vol. 5, no. 8, pp. 2841–2842 (2024).

[8] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Syst.,* vol. 226, p. 107134 (2021), doi: 10.1016/j.knosys.2021.107134.

[9] I. Raj, M. S. Jhala, and A. Patil, "Reviewing Flipkart product comments using methods based on sentiment analysis," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 4, no. 5, pp. 644–648 (2022).

[10] S. Kamış and D. Goularas, "Evaluation of deep learning techniques in sentiment analysis from Twitter data," in *Proc. IEEE,* pp. 12–17 (2019), doi: 10.1109/DeepML.2019.00011.

[11] A. Kumar, V. Tyagi, and S. Das, "Sentiment analysis on Twitter data using deep learning approach," in *Proc. IEEE,* pp. 187–190 (2021), doi: 10.1109/ICACCCN51052.2020.9362853.

[12] M. G. Dhote, B. M. Nanche, P. N. Mahalle, S. S. Ali, and V. S. Karwande, "Deep learning for optimized path planning in autonomous vehicles by integrating reinforcement learning with convolutional neural networks," *J. Inf. Optim. Sci.,* vol. 46, no. 4B, pp. 1129–1139 (2025), doi: 10.47974/JIOS-1897.

[13] R. K. Moje, B. Tiple, S. M. Patil, and A. Revekart, "A framework formulti-task learning optimization in deep neural networks: Balancing task priorities for improved performance," *J. Inf. Optim. Sci.,* vol. 46, no. 4B, pp. 1129–1139 (2025).

[14] A. R. Deshmukh, P. S. Dhumal, and S. Bhattacharya, "Adaptive noise injection techniques for optimizing deep learning models under adversarial attacks," *J. Inf. Optim. Sci.*, vol. 46, no. 4B, pp. 1153–1163 (2025).

Received April, 2025