IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

LITERATURE REVIEW ON: "SENTIMENT ANALYSIS IN E-COMMERCE USING BERT"

[1]Hrishikesh P. Bangar, [2]Prof. Manisha Desai, [3]Suyash Khedekar

[4]Aditya Daware, [5]Siddhesh Gabhane,

Department of Computer Engineering,

R.M.D. Sinhgad School of Engineering, Pune, India

Abstract: In the discipline of Natural Language Processing (NLP), sentiment analysis is an essential task that involves evaluating and analysing user opinion data on a given subject. Many deep learning models, including the recently announced Bidirectional Encoder Representations from Transformers (BERT) model, have been created to tacklethis difficulty. For the sentiment analysis task, we use datasets of Vietnamese reviews to test two BERT fine-tuning strategies in this study: Two methods are available: 1) one that feeds the [CLS] token alone into an associated feed-forward neural network; 2) another that uses all BERT output vectors as the input for classification. Experimental results on two datasets show that models utilising BERT perform models utilising Glove and Fast Text by a small margin.

Index Terms **– Sentiment Analysis, Deep Learning, Pre-trained Language Model, BERT.**

1. INTRODUCTION

In the rapidly evolving landscape of e-commerce, customer opinions and sentiments play a pivotal role in shaping the success of businesses. Understanding the sentiments expressed in user reviews, comments, and feedback can provide invaluable insights into customer satisfaction, preferences, and product perceptions. Sentiment analysis, a subfield of natural language processing (NLP), has emerged as a powerful tool to systematically evaluate and interpret these sentiments.

This project focuses on implementing sentiment analysis in the realm of e-commerce, leveraging the cutting-edge Bidirectional Encoder Representations from Transformers (BERT) model. BERT, a pre-trained transformer-based neural network architecture, has demonstrated remarkable success in capturing contextualized word embeddings and understanding the intricate nuances of natural language. By applying BERT to e-commerce datasets, we aim to develop a robust sentiment analysis model capable of classifying user sentiments into categories such as positive, negative, or neutral.

The primary objectives of this project include exploring two distinct fine-tuning strategies for BERT in the context of sentiment analysis. The first strategy involves feeding only the [CLS] token into a feed-forward neural network, while the second strategy utilizes all BERT output vectors for classification. The comparative analysis of these strategies will provide insights into the optimal approach for sentiment classification in the e-commerce domain.

As the e-commerce landscape continues to grow and diversify, understanding and responding to customer sentiments become crucial for businesses to stay competitive. This project seeks to contribute to the field of sentiment analysis by harnessing the capabilities of BERT, ultimately providing businesses with a powerful tool to gauge and respond to user sentiments in the dynamic world of online commerce.

2.LITERATURE SURVEY

In the realm of sentiment analysis and social media, several notable research contributions have been made, addressing diverse aspects of data analysis and classification. Bala Durga Dharmavarapu and Jayanag Bayana focus on sarcasm detection in Twitter, employing Naive Bayes and AdaBoost algorithms to discern subtle irony in tweets [3]. Zeel Doshi, Subhash Nadkarni, Kushal Ajmera, and Prof. Neepa Shah introduce "TweetAnalyzer," a real-time Twitter data extraction tool that aids in trend detection, hashtag popularity, and user activity visualization [8]. Arijit Chatterjee and Dr. William Perrizo present another facet of Twitter analysis with "TweetAnalyzer," emphasizing its versatility in applications such as job searching, news updates, and business intelligence [4]. Umit Demirbaga introduces "HTwitt," a Hadoop-based platform designed to handle big data challenges in classifying Twitter data. The framework employs algorithms and a Naïve Bayes classifier, showcasing its effectiveness in real-world scenarios such as early warning systems for landslides [6]. Ming Hao, Christian Rohrdantz, and Halldo´r Janetzko delve into visual sentiment analysis on Twitter data streams, proposing innovative techniques for topic-based sentiment analysis, stream analysis, and visualizations based on tweet density and sentiment [9]. Mehwish Rani and Seemab Latif conduct a survey on sentiment analysis techniques applied to textual data, demonstrating the effectiveness of transfer learning in analyzing sentiments in Pakistani YouTube user comments [1]. Pedro Faria presents a methodology for the aggregation and remuneration of electricity consumers and producers in the context of demand-response programs, essential for smart grid improvements [2]. Bharat Singh explores the challenges of online user-generated reviews across social media and e-commerce websites, highlighting the crucial role of opinions in decision-making for individuals and organizations [4] [10]. Finally, S. Uma Maheswari and S. S. Dhenakaran contribute to aspect-based sentiment analysis on social media big data, employing fuzzy logic and natural language processing to extract semantic aspects and opinions from customer tweets on platforms like Twitter and Flipkart [6].

3.ALGORITHMIC SURVEY:

Sr no.	Publication Detail	Algorithm used	Dataset	Accuracy	Research Gap Identified
1.	SPARK-BASED LARGE-SCALE MATRIX INVERSION FOR BIG DATA PROCESSING	Basic block-recursive matrix inversion algorithm	Kaggle	87%	optimizing the algorithm through reduced communication costs using emerging technologies like Tachyon
2.	Sentiment Analysis of Twitter Corpus Related to Artificial Intelligence Assistants	VADER	Kaggle	85%	absence of a detailed examination and comparison of machine learning-based sentiment analysis algorithms in evaluating user opinions on artificial intelligence assistants from Twitter data.

3.	Real Time Sentiment Analysis Of Twitter Posts	Bag of Words (BoW),	Kaggle	89%	The paper does not discuss the analysis of emoticons, which can convey sentiment. Including emoticon analysis could improve the accuracy of sentiment classification.
4.	Sentiment Analysis of Twitter Data in Online Social Network	supervised learning approach, specifically focusing on sentiment polarity calculation for each tweet.	Kaggle	90%	not addressing the impact or consideration of contextual factors in sentiment determination.
5.	Sentiment Analysis In Twitter Using Lexicon Based and Polarity Multiplication	lexicon-based sentiment analysis approach with NLP techniques	Kaggle	89%	incomplete lexicon and suggests further exploration into semantic analysis for polarity determination.
6.	AGGREGATION AND REMUNERATION OF ELECTRICITY Consumers and Producers for the Definition of Demand-Response Programs	hierarchical and fuzzy c-means clustering	Kaggle	87%	absence of a comprehensive evaluation of the clustering methods' performance and the need for further exploration into the impact of these tariff structures on overall smart grid efficiency.
7.	Sentiment Analysis for Arabic e-commerce websites	Naive Bayes and Support Vector Machine	Kaggle	89%	limited exploration of multi-source opinion extraction and the absence of a dedicated lexicon for Arabic sentiment expressions.

8.	DIRECTION DENSITY-BASED SECURE ROUTING PROTOCOL for Healthcare Data in Incompletely Predictable Networks	HMAC	Kaggle	84%	lack of comparison with existing secure routing protocols tailored for healthcare data transmission in incompletely predictable networks.
9.	ASPECT BASED FUZZY LOGIC SENTIMENT ANALYSIS on Social Media Big Data	Fuzzy Logic and Natural Language Processing	Kaggle	90%	lack of clarity on how the model handles diverse data types, such as images, audio, and video in customer reviews.
10.	A Survey of Sentiment Analysis of Internet Textual Data and Application to Pakistani YouTube User Comments	LSTM, GRU	Kaggle	87%	exploration of word embedding-based techniques, leaving a potential research gap.
11.	Issue and challenges of online user generated reviews across social media and e-commerce website	Opinion mining	Kaggle	85%	lack of sufficient resources and corpus for different languages used in various regions.
12.	BERT-IAN Model for Aspect-based sentiment analysis	BERT-IAN Model	Kaggle	90%	Adding location information of aspects and contexts can effectively improve the accuracy.
13.	BERT-based hierarchical model for Vietnamese aspect based sentiment analysis	BERT	Kaggle	90%	To improve the performance add other models like ELECTRA and GPT.

4.CONCLUSION & FUTURE SCOPE:

this review paper explores the significance of sentiment analysis in the domain of Natural Language Processing (NLP), with a particular focus on e-commerce. Leveraging the advanced Bidirectional Encoder Representations from Transformers (BERT) model, the study compares two fine-tuning strategies for sentiment analysis: one involving the [CLS] token and a feed-forward neural network, and another utilizing all BERT output vectors for classification. Experimental results on Vietnamese review datasets demonstrate that models employing BERT outperform those using Glove and Fast Text by a slight margin. The algorithmic survey presents a comprehensive overview of diverse sentiment analysis studies, highlighting the methodologies, algorithms, datasets, accuracies, and research gaps identified in each. These findings collectively contribute to advancing the field of sentiment analysis, emphasizing the effectiveness of BERT in capturing nuanced language patterns and providing valuable insights for businesses operating in the dynamic landscape of e-commerce.

The future scope for sentiment analysis using the BERT model includes optimizing fine-tuning strategies, exploring multilingual applications, and enhancing contextual analysis, presenting opportunities for more accurate and adaptable sentiment classification in diverse linguistic contexts and real-world scenarios.

5. REFERENCES

1. Anukarsh G Prasad, Sanjana S, Skanda M Bhat, B S Harish “Sentiment Anal-ysis for Sarcasm Detection on Streaming Short Text Data”, 2nd InternationalConference on Knowledge Engineering and Applications, IEEE, 2017.
2. Sana Parveen, Sachin N. Deshmukh, “Opinion Mining in Twitter – Sarcasm Detection” International Research Journal of Engineering and Technology (IR- JET), volume 04, issue 10, pages 201-204, October 2017.
3. Paras Dharwal, Tanupriya Choudary, Rajat Mittal, Praveen Kumar, “AutomaticSarcasm Detection using Feature Selection”, International Conference on Ap- plied and Theoretical Computing and Communication Technology, IEEE, 2017.
4. Sindhu. C, G. Vaidhu, Mandala Vishal Rao, “A Comprehensive Study on Sar- casm Detection Techniques in Sentiment Analysis”, International Journal of Pure and Applied Mathematics, volume 118, pages 433-442, 2018.
5. Tanya Jain, Nilesh Agrawal, Garima Goyal, Niyati Aggrawal, “Sarcasm Detection of Tweets: A Comparative Study”, Tenth International Conference on Contemporary Computing (IC3), IEEE, August 2017.
6. Levy, M. (2016). Playing with Twitter Data. [Blog] R-bloggers. Available at: https://www.r-

bloggers.com/playing-with-twitter-data/ [Accessed 7 Feb. 2018].

7. Popularity Analysis for Saudi Telecom Companies Based on Twitter Data. (2013). Research Journal of Applied Sciences, Engineering and Technology. [online] Available at: http://maxwellsci.com/print/rjaset/v6-4676-4680.pdf [Ac-cessed 1 Feb. 2018].

8. Zhao, Y. (2016). Twitter Data Analysis with R – Text Mining and Social Net- work Analysis. [online] University of Canberra, p.40. Available at:https://paulvanderlaken.files.wordpress.com/2017/08/rdataminingslides-twitter-analysis.pdf [Accessed 7 Feb. 2018].

9. Alrubaiee, H., Qiu, R., Alomar, K. and Li, D. (2016). Sentiment Analysis ofArabic Tweets in e-Learning. Journal of Computer Science. [online] Available at: http://thescipub.com/PDF/jcssp.2016.553.563.pdf [Accessed 7 Feb. 2018].

10. Qamar, A., Alsuhibany, S. and Ahmed, S. (2017). Sentiment Classification of Twitter Data Belonging to Saudi Arabian Telecommunication Companies. (IJACSA) International Journal of Advanced Computer Science and Applica- tions, [online] 8. Available https://thesai.org/Downloads/Volume8No1/Paper 50- Sentiment Classification of Twitter Data Belonging.pdf [Accessed 1 Feb.2018].

11. R. M. Duwairi and I.Qarqaz, “A framework for Arabic sentiment analysis us- ing supervised classification” , Int. J. Data Mining, Modelling and Manage- ment, Vol. 8, No. 4, pp.369-381 , 2016.

12. Hossam S. Ibrahim, Sherif M. Abdou, Mervat Gheith, “Sentiment Analysis For Modern Standard Arabic And Colloquial”, International Journal on Natu-ral Language Computing (IJNLC), Vol. 4, No.2, pp. 95-109, April 2015.