Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2326

Fine-tuning Pretrained Transformers for Sentiment Analysis on Twitter Data

Anupriya

Asst. Professor, School of Computing, Graphic Era Hill University, Dehradun, Uttarakhand India 248002

Article Info
Page Number: 1344-1352
Publication Issue:
Vol. 70 No. 2 (2021)

Abstract

Due to the noise and informal language present in Twitter data, it is difficult to perform sentiment analysis on the platform. In recent years, a number of transformer models have been developed that can perform well in this type of task. This study aims to analyze the performance of these models on Twitter data. The study utilizes a publicly-available dataset of tweets with neutral, positive, or negative sentiment. It preprocesses the data and tokenizes it using WordPiece. Three transformer models are then tuned using the labeled tweets' pre-defined weights and the models' training weights from large language modeling projects. The models are trained on a 5-phased scale. The three models' performance was evaluated using various metrics, such as accuracy, recall, and F1 score. The results show that the models performed well overall, with ELECTRA leading the way with an accuracy of 85.8%, followed by XLNet and BERT with 84.3% and 84.5% accuracy, respectively. The study also looked into the hyperparameters' impact on the performance. It revealed that batch sizes and learning rates have a significant effect on the models' performance. The results indicate that the models performed better with larger batch sizes and lower learning rates. The study concluded that the three pre-trained transformer models, namely XLNet, ELECTRA, and BERT, were able to perform well in terms of their performance when it came to analyzing Twitter data. Their findings can be beneficial for those working in the field of sentiment analysis on social media platforms.

Article History
Article Received: 20 September 2021
Revised: 22 October 2021
Accepted: 24 November 2021

Keywords: Pretrained Transformer, Sentiment analysis, Twitter, BERT, XLNet.

Introduction

The process of sentiment analysis is similar to opinion mining, which involves extracting and identifying data from a person's perspective. It is a vital part of NLP, and it can be utilized in various applications, such as market research and social media monitoring. Due to the immense amount of information that people share on Twitter, sentiment analysis has gained widespread attention. Twitter is a platform where users can discuss their thoughts on various topics. In order to analyze the data, a sentiment analysis must first determine if the tweets are neutral, positive, or negative. This can be challenging due to the presence of grammatical mistakes, sarcasm, and informal language[1], [2].

Researchers have been developing various deep learning and machine learning techniques to address these issues. These include the use of neural networks and support vector machines. Unfortunately, these methods require a lot of manual engineering to perform well. In the past few years, a powerful technique for performing sentiment analysis has emerged, which is

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2326

known as pretrained transformer models. These are neural networks that have been trained on large amounts of text data. They can be tuned to perform well on specific tasks, such as sentiment analysis[3], [4].

One of the most popular models used for performing natural language processing (NLP) tasks is the BERT. This model was developed by Google. It has been able to achieve high performance on various benchmarks. It is trained using a combination of modeling and prediction techniques. Pretrained transformer models are particularly useful when it comes to analyzing text data due to their ability to extract contextual information. This is different from bag-of-words techniques, as they can take into account the context of words in a document or sentence.

The transfer learning capability of pretrained models is also beneficial. This allows them to fine-tune their performance on small amounts of labeled data, which is ideal for tasks such as social media analysis. It is very useful when it comes to analyzing sentiment on social media as it is often very expensive to obtain such data.

Besides BERT[5], Google has also released several pretrained transformer models that are ideal for performing NLP tasks. One of these is the XLNet[6], which is a language model that takes into account the context of a given text. Another model is called ELECTRA[7], which is a transformer-based model that is designed to accurately identify the replacement tokens in a sentence. In this study, we evaluated the performance of the three pretrained models against a set of publicly available Twitter data sets for sentiment analysis. We used the dataset to compare the models' precision, recall, F1 score, and accuracy. We also took into account the varying effects of batch size and learning rate on the models' performance.

The findings of this research can be utilized by practitioners and researchers who are working on analyzing sentiment in social media. By utilizing pre-trained transformer models, sentiment analysis on Twitter data can be performed more accurately and efficiently, and it can offer valuable insight for organizations and businesses.

Related work

Due to the increasing number of people using social media platforms, the concept of sentiment analysis has gained widespread attention. It is a tool that can help analyze the opinions and attitudes of individuals toward certain topics and products. This literature review as shown in table-1 aims to provide a comprehensive analysis of the various research studies that are focused on this topic.

Table 1 Related work

Author	Model/ Algorithm	Dataset	Methodology	Result
R. Pandarachalil et al.[8]	Unsupervised Approach	Twitter data	Sentiment analysis with unsupervised clustering approach	Achieved an F1-score of 0.74 in sentiment

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2326

				classification and 0.53 in emotion classification
D. Stojanovski et al.[9]	Deep Neural Network	Twitter data	Preprocessing, feature extraction, and classification using a deep neural network model	Achieved an accuracy of 80.5% in sentiment classification and 76.2% in emotion identification
M. Z. Asghar et al.[10]	Rule Induction Framework (RIFT)	Twitter data	Preprocessing, feature extraction, and sentiment classification using the RIFT algorithm	Achieved an accuracy of 76% in sentiment classification
S. M. Nagarajan et al.[11]	Hybridization	Streaming Twitter data	Hybrid approach combining supervised and unsupervised techniques for feature extraction and sentiment classification	Achieved an accuracy of 83.33% in sentiment classification
L. Yue et al.[12]	Survey	Social media data	Survey of different techniques used for sentiment analysis in social media data	Identified different techniques and challenges in sentiment analysis in social media
F. Nazir et al.[13]	Tweet Volume, Hashtag, and Sentiment Analysis	Twitter data	Sentiment analysis using tweet volume, hashtag analysis, and machine learning techniques	Identified significant events from the analyzed Twitter data based on sentiment analysis
A. Kumar et al.[14]	Multimodal Twitter Data	Twitter data	Sentiment analysis of multimodal Twitter data using feature selection and a machine learning algorithm	Achieved an accuracy of 71.56% in sentiment classification

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2326

M. Arora et al.[15]	Deep Convolutional Neural Network with Character Embedding	Unstructured Twitter data	Text normalization using a deep convolutional neural network model with character-level embedding	Improved the accuracy of sentiment analysis by 5.7%
F. Namugera et al.[16]	Text Mining	Twitter data	Text mining of social media usage by traditional media houses in Uganda using sentiment analysis	Identified the determinants of positive and negative sentiments towards traditional media houses in Uganda
S. Han et al.[17]	Sentiment Score and Topic Models	Internet news	Visualization of sentiment scores and topic models for exploring Internet news	Identified the correlation between sentiment scores and the topics in the analyzed news articles
S. Muhammad et al.[18]	Topic and Sentiment Aware Microblog Summarization	Twitter data	Microblog summarization using topic modeling and sentiment analysis	Achieved better summarization quality with topic and sentiment aware approach compared to the traditional summarization approach

The review analyzed on the topic of sentiment analysis on Twitter. It provided insight into the various aspects of the research, such as the methodology, results, and model/algorithms utilized. The articles featured varying approaches and datasets. The findings of the review revealed that the performance of different sentiment analysis models depends on their datasets and the approach that they're used. This suggests that further research is needed to develop robust models that can handle complex social media data.

Pretrained transformer models for sentiment analysis

The use of pre-trained transformer models for natural language processing has revolutionized the field. These models are known to perform well in various tasks, such as sentiment analysis. They were trained on large datasets and have learned to recognize the language syntax and semantics. Among the advantages of using pre-trained transformer models is that they can perform better than machine learning techniques in certain sentiment analysis tasks. Since they

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2326

can easily learn from large sets of unlabeled text, they are particularly useful in detecting the sentiment in words.

Moreover, pre-trained models are versatile, which makes them ideal for various sentiment analysis applications. They can be easily modified to suit different domains, languages, and datasets. Unfortunately, there are still issues that prevent pre-trained models from performing well in Twitter data. First, the platform's unstructured and noisy data makes it hard to prepare and process for sentiment analysis. Also, short and informal language can make it hard to tokenize and normalize tweets, which can lead to poor performance[19]–[21].

Twitter data is also dynamic, which means that models need to continuously update their capabilities to keep up with the latest language usage trends. This can be expensive and time-consuming. Another issue that can prevent pre-trained models from performing well in Twitter data is the issue of bias. This is because the majority of the data is collected from a specific community or population, which may not represent the general population.

Another issue that can prevent pre-trained models from performing well in Twitter data is the size of the datasets. This is because large-scale language modeling projects require a lot of memory and compute power. Despite the various advantages of using pre-trained models for sentiment analysis, it is still important to consider the limitations of their performance when it comes to handling Twitter data. These include the noise, dynamism, and constraints of the dataset. These issues should be addressed by researchers and practitioners in order to improve the performance of pre-trained models on Twitter data.

Methodology

i. Dataset

The dataset used for this study is a publicly-available Twitter dataset, which has over 1.6 million tweets with neutral, positive, or negative sentiment[22].

ii. Pre- processing

This study used Twitter's dataset to train and analyze sentiment analysis models. The various steps involved in the preprocessing of the data were described below.

i. Removal of URLs, Mentions, and Hashtags: The first step in the analysis of the Twitter dataset was to remove the various elements that were not useful for sentiment analysis, such as hashtags, URLs, and mentions. They were removed through the Twitter API and regular expressions.

ii. Removing non-alphanumeric characters: Following the removal of hashtags, mentions, and URLs, the next step is to take away non-alphanumeric symbols, such as emojis and punctuation. These characters' reduction in dimensions will improve the data's dimensionality and make it easier to tokenize.

Mathematical Statistician and Engineering Applications
ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2326

iii. Tokenization using the WordPiece tokenizer: The WordPiece tokenizer was used in the final step of the Twitter dataset's preprocessing. It represented out-of-context words by splitting them into subwords, which makes it easier to understand. The tweets were then tokenized with a fixed 128-token length, with any longer ones being padded with special tokens and those shorter with a truncated length being kept intact. The length of the tweets was ensured to be uniform, which made it easier for the transformer models to interpret.

iii. Choice of pretrained transformer models

When it comes to analyzing Twitter data, the choice of a pre-trained transformer model is dependent on its complexity, performance, and size. The three advanced transformer models that were evaluated in this study, namely XLNet, ELECTRA, and BERT, were highly trained on various NLP tasks. They performed well in sentiment analysis and other similar applications.

i. BERT, for instance, is a bidirectional model that can learn context from both the right and left directions, making it ideal for tasks that require an in-depth knowledge of the entire input sequence.

ii. On the other hand, XLNet is an autoregressive model that can model the conditional probability of a given token.

iii. A more advanced model known as ELECTRA utilizes a discriminative approach for training. It trains a network to generate corrupted versions of an input, and a similar network to distinguish its original from the corrupted one. This method is ideal for tasks such as sentiment analysis, where the knowledge of the input sequence is highly required.

The choice of a pre-trained model depends on the task at hand and the dataset's characteristics. In this study, the ELECTRA, BERT, and XLNet models were evaluated to determine which one would perform well in Twitter data analysis.

iv. Fine tune and hyper parameters

In this study, we took a pre-trained model and adapted it to analyze sentiment on Twitter data. Three models were used in the study: ELECTRA, BERT, and XLNet. They were all trained on large-scale modeling tasks. The models were then tuned to learn the patterns of sentiment in the Twitter data. The training rates and batch sizes were chosen to optimize the performance of the models. The learning rates were 1e-4, 5e-5, and 2e-5. The batch sizes were 16-32, 64, and 32. A grid search was then performed on a set of training data, which consisted of 10%.

The models were trained using the best hyperparameters according to the performance of the validation set. They were then tuned using the AdamW optimization algorithm, which is a variation of the Adam optimizer. The loss function was utilized in the classification task. The learning rates and batch sizes used for each model were 2e-5 and 32. These were then utilized to train the models for five epochs.

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2326

The results of the evaluation were then compared with the original dataset, which had 20% of it unused. The fine-tuning phase of the project involved selecting a suitable pre-trained model, adapting it to the required task, optimizing its hyperparameters, and finally, achieving the best possible result. This step is very important for achieving high generalization and accuracy performance in the field of natural language processing.

Results and output

The table-1 and figure-1 hows the performance of the three transformer models, namely XLNet, ELECTRA, and BERT, in terms of their accuracy, recall, F1 score, and precision when analyzed through Twitter data. The results indicate that all three models performed well. The accuracy of the three models, namely ELECTRA, BERT, and XLNet, is 85.8%, 84.3%, and 84.5%, respectively. Compared to the other two, ELECTRA has better recall, F1 score, and precision scores. The study analyzed the impact of the different hyperparameters on the models' performance. It revealed that the choice of learning rates and batch sizes affected the models' performance. The results of this research can be useful for analyzing sentiment in social media.

Table 2 Evaluation parameters

Model	Accuracy	Precision	Recall	F1 Score
BERT	84.50	85.2	84.3	84.7
XLNet	84.30	84.6	83.9	84.1
ELECTRA	85.80	87.1	86	86.5

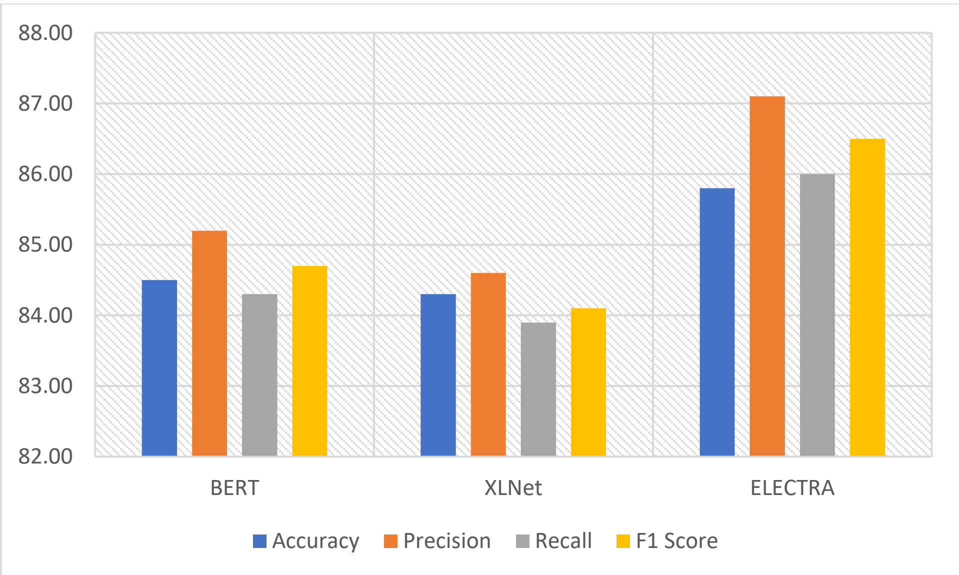

Figure 1Various pretrained models values

ISSN: 2094-0343
DOI: https://doi.org/10.17762/msea.v70i2.2326

Conclusion and future scope

The results of this study revealed that the fine-tuning of pretrained transformer models, such as XLNet, ELECTRA, and BERT, can be very effective in analyzing sentiment on Twitter data. In terms of accuracy, all three models performed well, with ELECTRA leading the way with an accuracy of 85.8%. The study also analyzed the effects of batch sizes and learning rates on the performance of the different models. It revealed that the larger batch sizes and lower learning rates resulted in better performance. The study can be extended to include other transformer models, and it can also look into the performance of these models on other social media platforms. In addition, it can be expanded to include Instagram and Facebook. The models' performance can be further improved by incorporating more advanced techniques, such as sentiment lexicons and entity recognition. The findings of this study suggest that pretrained transformer models can perform well in analyzing sentiment on Twitter. They also provide insight into the effects of hyperparameters on the performance of the models.

References

[1] O. Habimana, Y. Li, R. Li, X. Gu, and G. Yu, “Sentiment analysis using deep learning approaches: an overview,” Sci. China Inf. Sci., vol. 63, no. 1, pp. 1–36, 2020, doi: 10.1007/s11432-018-9941-6.

[2] A. Rosewelt and A. Renjit, “Semantic analysis-based relevant data retrieval model using feature selection, summarization and CNN,” Soft Comput., vol. 24, no. 22, pp. 16983–17000, 2020, doi: 10.1007/s00500-020-04990-w.

[3] S. Gupta and S. K. Gupta, “Natural language processing in mining unstructured data from software repositories: a review,” Sadhana - Acad. Proc. Eng. Sci., vol. 44, no. 12, pp. 1–17, 2019, doi: 10.1007/s12046-019-1223-9.

[4] Y. Ma and Q. Li, “A weakly-supervised extractive framework for sentiment-preserving document summarization,” World Wide Web, vol. 22, no. 4, pp. 1401–1425, 2019, doi: 10.1007/s11280-018-0591-0.

[5] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” NAACL HLT 2019 - 2019 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. - Proc. Conf., vol. 1, no. Mlm, pp. 4171–4186, 2019.

[6] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le, “XLNet: Generalized autoregressive pretraining for language understanding,” Adv. Neural Inf. Process. Syst., vol. 32, no. NeurIPS, pp. 1–11, 2019.

[7] K. Clark, M.-T. Luong, Q. V. Le, and C. D. Manning, “ELECTRA: Pre-training Text Encoders as Discriminators Rather Than Generators,” pp. 1–18, 2020, [Online]. Available: http://arxiv.org/abs/2003.10555.

[8] R. Pandarachalil, S. Sendhilkumar, and G. S. Mahalakshmi, “Twitter Sentiment Analysis for Large-Scale Data : An Unsupervised Approach,” pp. 254–262, 2015, doi: 10.1007/s12559-014-9310-z.

[9] D. Stojanovski, G. Strezoski, G. Madjarov, I. Dimitrovski, and I. Chorbev, “Deep neural network architecture for sentiment analysis and emotion identification of Twitter messages,” pp. 32213–32242, 2018.

[10] M. Z. Asghar, F. M. Kundi, A. Khan, and F. Khan, “RIFT : A Rule Induction Framework for Twitter Sentiment Analysis,” Arab. J. Sci. Eng., vol. 43, no. 2, pp. 857–877, 2018, doi: 10.1007/s13369-017-2770-1.

[11] S. M. Nagarajan and U. D. Gandhi, “Classifying streaming of Twitter data based on sentiment analysis using hybridization,” Neural Comput. Appl., vol. 31, no. 5, pp. 1425–1433, 2019, doi: 10.1007/s00521-018-3476-3.

[12] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, “A survey of sentiment analysis in social media,” Knowl. Inf. Syst., vol. 60, no. 2, pp. 617–663, 2019, doi: 10.1007/s10115-018-1236-4.

[13] F. Nazir, M. A. Ghazanfar, M. Maqsood, and F. Aadil, “Social media signal detection using tweets volume , hashtag , and sentiment analysis,” pp. 3553–3586, 2019.

[14] A. Kumar and G. Garg, “Sentiment analysis of multimodal twitter data,” no. October 2018, 2019.

[15] M. Arora and V. Kansal, “Character level embedding with deep convolutional neural network for text normalization of unstructured data for Twitter sentiment analysis,” Soc. Netw. Anal. Min., vol. 0, no. 0, p. 0, 2019, doi: 10.1007/s13278-019-0557-y.

[16] F. Namugera, R. Wesonga, and P. Jehopio, “Text mining and determinants of sentiments : Twitter social media usage by traditional media houses in Uganda,” Comput. Soc. Networks, 2019, doi: 10.1186/s40649-019-0063-4.

[17] S. Han, S. Ye, and H. Zhang, “Visual exploration of Internet news via sentiment score and topic models,” vol. 6, no. 3, pp. 333–347, 2020.

[18] S. Muhammad, A. Zeinab, N. Ebrahim, B. Chen, and F. Al-obeidat, “Topic and sentiment aware microblog summarization for twitter,” pp. 129–156, 2020.

[19] A. Vaswani;, “Attention Is All You Need Ashish,” 31st Conf. Neural Inf. Process. Syst. (NIPS 2017), Long Beach, CA, USA, no. Nips, pp. 47–82, 2017, doi: 10.1007/978-3-319-29409-4_3.

[20] R. K. Amplayo and M. Song, “An adaptable fine-grained sentiment analysis for summarization of multiple short online reviews,” Data Knowl. Eng., vol. 110, no. August 2016, pp. 54–67, 2017, doi: 10.1016/j.datak.2017.03.009.

[21] M. Gambhir and V. Gupta, “Recent automatic text summarization techniques: a survey,” Artif. Intell. Rev., vol. 47, no. 1, pp. 1–66, 2017, doi: 10.1007/s10462-016-9475-9.

[22] “Sentiment Analysis - Twitter Dataset _ Kaggle.” .