Turkish Online Journal of Qualitative Inquiry (TOJQI)
Volume 11, Issue 3 (ISSN:1309-6591) September 2020: 1052 - 1058
DOI: 10.53555/tojqi.v11i3.10629

Research Article

Fine-Tuning Bert For Financial Sentiment Analysis

Dr. Kiran Bala[1*]

Abstract

Sentiment analysis is a prominent research area in Natural Language Processing. It involves deducing a statement's underlying meaning, context, purpose, and emotional tone using NLP techniques. Sentiment analysis has been especially desired in the finance industry to draw plausible conclusions from the huge array of daily financial news and make satisfactory predictions about the market. Several models such as Roberta, XLNet, and BERT have previously been used for this task, however, they require a significantly large training corpus and training time. In this paper, we propose a BERT model fine-tuned on financial text corpus significantly smaller than solutions proposed in previous literature and achieve accuracy on par with other models.

I.INTRODUCTION

In recent years, sentiment analysis has emerged as a critical research area within Natural Language Processing (NLP). It involves the extraction of underlying meaning, context, purpose, and emotional tone from textual data, enabling a deeper understanding of the sentiments expressed within the text. This capability has found numerous applications across various industries, with the finance sector being of particular interest. The finance industry seeks to leverage sentiment analysis to extract valuable insights from the vast amount of daily financial news and, in turn, make informed predictions about the market.

In the pursuit of efficient and accurate sentiment analysis models for financial text, several NLP techniques and models have been proposed. Notably, the models Roberta, XLNet, and BERT have demonstrated remarkable performance in sentiment analysis tasks. However, a common challenge with these models lies in their reliance on extensive training data and substantial training time. The requirement of a large training corpus can be a significant barrier in real-world scenarios, where resources may be limited, and time is of the essence. To address these limitations and pave the way for more accessible and time-efficient sentiment analysis in finance, this paper introduces a novel approach. We propose a BERT-based sentiment analysis model fine-tuned on a financial text corpus significantly smaller than those used in previous literature. The primary objective of this work is to demonstrate that even with a reduced training corpus, it is possible to achieve comparable accuracy to existing models.

Throughout the remainder of this paper, we will delve into the details of our proposed methodology. Section II will provide an overview of related works in sentiment analysis, highlighting the significance of existing techniques and models in the finance domain. In Section III, we will present the BERT model and its underlying architecture, as well as discuss its advantages and limitations. Section IV will elaborate on the design and construction of our reduced financial text corpus for fine-tuning the BERT model. In section V, we describe the training process of our model. We will then delve into the specifics of the fine-tuning process in Section VI, explaining the modifications and optimizations made to adapt BERT for sentiment analysis in finance. The researchers of BERT have also added that fine-tuning has little effect on larger datasets (over 100K+ labeled training examples) as compared to smaller datasets. Hence, for our context, fine-tuning plays a crucial role in enhancing model performance.

[1] Assistant Professor, Department of Commerce, Delhi School of Economics, University of Delhi

Section VII will be dedicated to the evaluation of our proposed model, where we will present the experimental setup, performance metrics, and comparative analysis against other well-known sentiment analysis models. Furthermore, we will discuss the implications of our results and provide insights into the feasibility of employing a smaller training corpus for similar tasks.

Finally, we will conclude our findings in Section VIII, summarizing the contributions of this research and discussing potential areas for future improvement. By the end of this paper, readers will gain a comprehensive understanding of our novel BERT-based sentiment analysis model, its capabilities, and its potential impact on the finance industry.

II.LITERATURE SURVEY

The authors of [1] present a domain-specific BERT model pre-trained on financial communication text. Their model, FinBERT outperforms generic BERT models in financial sentiment classification tasks, offering a valuable resource for researchers and practitioners in financial NLP.

[2] explores the use of robo-readers for analyzing financial news texts, focusing on detecting semantic orientations in financial and economic news. The authors introduce a human-annotated finance phrase-bank as a benchmark, enhance financial lexicons with attributes to identify the expected direction of events, and develop a linearized phrase-structure model. Through a comparative study, they demonstrate the significance of their approach against previous general sentiment models and word frequency models, highlighting the potential benefits in computational finance and sentiment analysis.

Combining symbolic and subsymbolic methods is a promising strategy for addressing AI's growing complexity, notably in Targeted Aspect-based Financial Sentiment Analysis (TABFSA). The authors of [3] propose anterior, parallel, and posterior knowledge integration to leverage external lexical knowledge effectively. Experiments on FiQA Task 1 and SemEval 2017 Task 5 datasets demonstrate knowledge-enabled models outperform plain deep learning counterparts, some even surpassing state-of-the-art results in aspect sentiment analysis error. Parallel knowledge integration is identified as the most effective approach, and domain-specific lexical knowledge proves crucial in their ablation analysis.

The study [4] explores the use of contextual word embeddings, specifically the XLNet language model, for sentiment analysis on cryptocurrency news in both English and Malay languages. Various state-of-the-art pre-trained language models like BERT, ALBERT, ELECTRA, RoBERTa, and XLNet are compared for text representation. The XLNet-GRU model, which combines XLNet with Gated Recurrent Unit (GRU), is evaluated against other pre-trained language models. Utilizing manually labeled corpora of English and Malay news in the cryptocurrency domain, the XLNet-GRU sentiment regression model demonstrated superior performance over lexicon-based baselines, achieving mean adjusted R2 values of 0.631 for English and 0.514 for Malay across Bitcoin and Ethereum news datasets.

The authors of [5] propose a BERT-based approach for sentiment analysis and key entity detection in online financial text mining. Their method outperforms SVM, LR, NBM, and BERT in two financial sentiment analysis and key entity detection datasets, demonstrating its effectiveness in extracting crucial information from massive negative financial texts.

The dissertation [6] delves into sentiment analysis, also known as opinion mining or emotional mining, which aims to identify sentiments expressed in text and written data. With the rise of online commerce data and social media platforms like Twitter, sentiment analysis has become crucial for businesses and organizations. The research focuses on performing sentiment analysis using transformer-based pre-trained models like BERT and XLNet, as traditional machine and deep learning models struggle with fewer training instances and extensive feature engineering. The objective is to study the performance of BERT and XLNet with fewer training instances using Mixout regularization for stock market sentiment analysis. The proposed model shows improved performance in accuracy, precision, recall, and F1-score for both BERT and XLNet when given adequate and under-sampled data.

Comparison and fine-tuning of methods for Financial Sentiment
Financial sentiment analysis faces challenges due to limited labeled datasets and domain-specific language. This research reviews successful approaches, evaluating them on financial news headlines and stock market-related tweets. Extensive experiments on BERT and its variations achieve 87.1% accuracy on the financial PhraseBank dataset, surpassing previous methods. The study demonstrates BERT's potential in financial sentiment analysis, showcasing its superior accuracy through fine-tuning and logistic regression.

III.PROPOSED MODEL

1. Model Architecture: The BERT-based classifier is implemented as a neural network in PyTorch, utilizing the nn.Module class. The core components of the model are as follows:

a) BERT Model: The backbone of our sentiment analysis model is the BERT language representation model. We employ the BertModel from the Hugging Face's transformers library and initialize it with the pre-trained 'bert-base-cased' version. This base BERT model is capable of encoding contextual information from the input text and has demonstrated exceptional performance in a wide range of NLP tasks.

b) Dropout Layer: To prevent overfitting and improve generalization, we incorporate a dropout layer after the BERT model. The nn.Dropout module with a dropout probability of 0.5 is utilized to randomly deactivate neurons during training, which helps to regularize the model.

c) Output Layer: The final layer of the classifier is a linear layer, nn.Linear, with an output size corresponding to the number of sentiment classes (e.g., positive, negative, neutral). This layer takes the pooled output from the BERT model and maps it to the sentiment class probabilities.

2. Model Forward Pass: The forward pass of our BERT-based classifier involves the following steps:
a) Input Encoding: During the forward pass, the input text is tokenized and converted into input IDs and attention masks. The input IDs represent the indices of the tokens in the BERT vocabulary, while the attention mask is used to indicate which tokens are actual words and which are padding tokens.

b) BERT Processing: The input IDs and attention masks are passed to the BERT model, and the BertModel processes the tokens to produce the last hidden state and a pooled output. The last hidden state contains the contextual embeddings of each token, while the pooled output represents the aggregated representation of the entire input text.

c) Dropout Layer: The pooled output from BERT is passed through the dropout layer, randomly deactivating neurons to reduce overfitting and enhance model generalization.

d) Output Layer: Finally, the output of the dropout layer is fed into the linear output layer. The linear layer applies a linear transformation to the pooled output, mapping it to sentiment class probabilities.

IV.DATASET

The dataset curated for training the BERT model is a combination of two distinct datasets, namely the "Financial Phrasebank" dataset and the "Financial Question Answering (FiQA)" dataset. The amalgamation of these two datasets enables us to create a comprehensive and diverse corpus for training our sentiment analysis model.

1. Financial Phrasebank Dataset:

The Financial Phrasebank dataset comprises 4845 unique text-sentiment pairs. This dataset is a specialized collection of financial phrases and statements annotated with sentiment labels, including positive, negative, and neutral sentiments. It was chosen for its domain-specific nature, focusing on financial-related language, and its suitability for training a sentiment analysis model tailored to the finance industry. The phrases and statements in this dataset encompass a wide range of financial topics, making it highly relevant for sentiment analysis in the context of the stock market, investment decisions, and overall economic trends.

2. FiQA Dataset:

The FiQA Dataset, an open challenge dataset for financial sentiment analysis, is an invaluable resource comprising 1,111 text sentences. The dataset presents a unique task wherein the objective is to predict the associated numeric sentiment score for each English text sentence in the financial domain. The sentiment scores ranged from -1 to 1, covering a wide spectrum of sentiments, from highly negative to strongly positive.

Upon merging the Financial Phrasebank dataset and the FiQA dataset, our curated training corpus consists of 5957 unique text-sentiment pairs. These diverse and domain-specific text samples provide a broad coverage of financial language, including discussions on stocks, companies, financial news, and expert opinions. The fusion of these datasets ensures that the BERT model is well-equipped to handle a wide array of financial language variations and sentiments, enabling it to deliver robust and accurate sentiment analysis results within the finance domain.

The selection of both the Financial Phrasebank dataset and the FiQA dataset was made based on their suitability for our specific research objective. The Financial Phrasebank dataset is well-established and widely used for sentiment analysis tasks in the finance industry, offering a rich assortment of financial phrases with sentiment annotations. On the other hand, the FiQA dataset complements this by introducing a different perspective with its focus on financial questions and answers, encapsulating real-world scenarios where sentiment analysis can be crucial in understanding investor sentiments and market trends.

The combination of these datasets caters to the unique requirements of our BERT-based sentiment analysis model, which aims to achieve high accuracy with a smaller training corpus. The diverse nature of the merged dataset ensures that the model generalizes well to various financial sentiment expressions and can effectively capture sentiment nuances in financial text, ultimately contributing to more accurate market predictions and investment decisions.

V.TRAINING

Training first begins with the mapping of all the possible sentiments in the dataset (negative, neutral, & positive) to nominal values. After splitting the dataset into train, test, and validation subsets, next, encode the text in each training example using a tokenizer. The tokenizer converts the text values into numerical representations, which can be processed by the BERT model.

1. Tokenization: Tokenization is a crucial preprocessing step that converts raw text into a sequence of numerical tokens. We utilize the BERT tokenizer, which is specifically designed for the BERT model architecture. The tokenizer segments the input text into subword tokens, such as words or smaller fragments, and maps each token to its corresponding index in the BERT vocabulary. Additionally, the tokenizer inserts special tokens, such as the [CLS] token at the beginning of each text and [SEP] token to separate different segments. These tokens serve as input indicators and aid the BERT model in understanding the sentence structure and context.

2. Padding and Truncation: BERT processes input sequences of fixed length. To ensure uniformity across the dataset, input sequences are padded or truncated as required. Padding involves adding special padding tokens [PAD] to the input sequences to achieve a consistent length, while truncation is applied to cut sequences that exceed the specified maximum length. This ensures that all input sequences have the same length and can be efficiently batched for training.

3. Input Encoding: With tokenization, padding, and truncation completed, each text in the dataset is now represented as a sequence of numerical token IDs. Simultaneously, attention masks are created, which are binary tensors indicating the positions of valid tokens (1 for valid tokens, 0 for padding tokens). These encoded inputs are then organized into batches, ready for training.

4. BERT Model Initialization: The BERT-based classifier model is initialized using the pre-trained 'bert-base-cased' version. The model comes with a pre-trained set of weights that capture a rich understanding of language from a large corpus of text. These pre-trained weights serve as the starting point for training our sentiment analysis model.

5. Parameter Updation: This process involves training the initialized BERT model on the labeled financial text data from the training subset. The model parameters are updated using backpropagation and stochastic gradient descent to minimize the cross-entropy loss between the predicted sentiment probabilities and the ground truth sentiment labels. The model learns to adapt its pre-trained representations to better capture the sentiment patterns specific to the financial domain.

6. Dropout Regularization: To improve the model's generalization and reduce overfitting, we employ dropout regularization. The dropout layer is included after the BERT model, which randomly deactivates neurons during training with a dropout probability of 0.5. This dropout mechanism prevents the model from becoming overly reliant on specific features and promotes robustness in predictions.

7. Validation: Throughout the training process, model performance is evaluated on the validation subset at regular intervals. Validation allows us to monitor the model's progress and identify potential overfitting. If the model's performance on the validation set does not improve or starts to degrade, early stopping can be implemented to prevent further training.

8. Testing and Performance Evaluation: Once the fine-tuning process is complete, the trained BERT-based classifier is evaluated on the test subset. Performance metrics such as accuracy, precision, recall, and F1-score are computed to assess the model's effectiveness in sentiment analysis for financial text. These metrics provide insights into the model's ability to correctly classify sentiment labels, thereby measuring the overall quality of predictions.

VI.FINE-TUNING & TESTING

The optimizer selected was AdamW, along with a Linear Scheduler with warmup enabled.

BERT-Base-Cased

Model	Hyperparameters	Accuracy	Precision (Weighted)	Recall (Weighted)	F1 Score (Weighted)	Loss	Time (in Seconds)
1.	lr = 2e-5 eps = 1e-5 betas = (0.9, 0.999) Batch_size = 8	0.879	0.879	0.879	0.879	0.428	2905.917
2.	lr = 3e-5	0.872	0.873	0.872	0.872	0.364	2745.55

	eps = 1e-8 Batch_size = 16						
3.	lr = 2e-5 eps = 1e-8 Batch_size = 16	0.869	0.873	0.869	0.868	0.826	2790.392

BERT-Base-Uncased

Model	Hyperparameters	Accuracy	Precision (Weighted)	Recall (Weighted)	F1 Score (Weighted)	Loss	Time (in Seconds)
1.	lr = 2e-5 eps = 1e-5 betas = (0.9, 0.999) Batch_size = 8	0.862	0.866	0.862	0.863	0.372	2970.976
2.	lr = 2e-5 eps = 1e-8 Batch_size = 16	0.869	0.873	0.869	0.868	0.826	2790.392

VII.COMPARATIVE STUDY/RESULTS

	Our Model	FinBERT (PhraseBank)	FinBERT (FiQA)	XLNet-GRU
Accuracy	0.899	0.872	0.844	-
Adjusted R^2		-	-	0.654

VIII.CONCLUSION

Our fine-tuned BERT-based sentiment analysis model has demonstrated appreciable capabilities in providing accurate and insightful sentiment predictions for financial text. Through the amalgamation of the Financial Phrasebank and FiQA datasets, we curated a diverse and domain-specific training corpus, allowing our model to effectively capture the intricacies of sentiment expressions within the finance industry. The BERT model's contextual understanding and fine-tuning of the reduced financial text corpus have proven to be highly advantageous, as our model achieved competitive accuracy with significantly less training data compared to existing models.

IX.ACKNOWLEDGMENT

X.REFERENCES

1. Y. Yang, M. C. S. UY, and A. Huang, “FinBERT: A Pretrained Language Model for Financial Communications.” 2020.
2. P. Malo, A. Sinha, P. Takala, P. Korhonen, and J. Wallenius, “Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts.” 2013.

3. K. Du, F. Xing, and E. Cambria, “Incorporating Multiple Knowledge Sources for Targeted Aspect-Based Financial Sentiment Analysis,” *ACM Trans Manage Inf Syst*, vol. 14, no. 3, Jun. 2023, doi: 10.1145/3580480.
4. N. A. Mohamad Zamani, J. S. Y. Liew, and A. M. Yusof, “XLNET-GRU Sentiment Regression Model for Cryptocurrency News in English and Malay,” in *Proceedings of the 4th Financial Narrative Processing Workshop @LREC2022*, Marseille, France: European Language Resources Association, Jun. 2022, pp. 36–42. [Online]. Available: https://aclanthology.org/2022.fnp-1.5
5. L. Zhao, L. Li, and X. Zheng, “A BERT based Sentiment Analysis and Key Entity Detection Approach for Online Financial Texts.” 2020.
6. Jangir, S. (2021). Finetuning BERT and XLNet for Sentiment Analysis of Stock Market Tweets using Mixout and Dropout Regularization. Technological University Dublin.