ChatGPT Tweets Sentiment Analysis

Sravanthi Garuda

Abstract

Sentiment Analysis plays a vital role in natural language processing by deciphering opinions, attitudes, and emotions embedded in textual data, enabling the determination of an overall sentiment or emotional tone. This project delves into the realm of sentiment analysis, utilizing a Twitter dataset as a rich source of diverse textual data, presenting an ideal opportunity for sentiment analysis.

The project involves a pre-processing pipeline, encompassing tokenization, cleansing, and vectorization of raw text data. Subsequently, deep learning models, including Long Short-Term Memory (LSTM), Bidirectional LSTM, and XLM models are trained on the preprocessed data. These models are designed to grasp intricate patterns and relationships between words and sentiments, leveraging a labeled dataset of tweets annotated with sentiment labels.

A distinctive aspect of this study lies in its comprehensive analysis of the performance, advantages, and limitations of the employed methods. The evaluation is conducted under a unified testing framework, maintaining consistency through the utilization of the same dataset and computing environment. The findings contribute valuable insights of the complex models in the context of sentiment analysis, advancing our understanding of their practical applications and implications.

A branch of artificial intelligence called natural language processing (NLP) studies how human languages and computers interact. It includes the creation of models and algorithms that let computers comprehend, interpret, and produce language that is similar to that of humans. The development of deep learning, a branch of machine learning, has transformed natural language processing (NLP) in recent years, enabling more complex and context-aware language processing. The use of Deep learning is the art of extracting complex patterns and representations from data through the use of multi-layered artificial neural networks, or deep neural networks. This technology has shown impressive results in several fields, such as speech recognition, computer vision, and NLP in particular.

One popular application of NLP, which stands for natural language processing, is sentiment analysis, which is the process of extracting and interpreting attitudes, opinions, and emotions from textual data. Determining the general sentiment—whether favorable, negative, or neutral—expressed in a text is the aim. This work is essential to corporate intelligence, customer relationship management, and social media monitoring since it helps to comprehend public opinion and consumer feedback, and helps in the future scope for developments of the IT industry.

Sentiment analysis is significant because it can extract meaningful information from large volumes of unstructured text. Companies can monitor client happiness, modify their marketing plans, and respond to problems instantly. Sentiment analysis is a tool used by social media networks to identify patterns, foresee possible problems, and improve user experience. Additionally, it supports the tracking of public opinion regarding social issues, political developments, and goods, which helps with well-informed decision-making.

We explore the complex nexus of sentiment analysis, deep learning, and natural language processing in this research. We investigate the efficiency of three distinct deep learning models Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Bidirectional LSTM—in identifying attitudes, using a Twitter dataset as our main source of textual data. Our goal is to comprehend the complex performance, advantages, and constraints of various models under a single testing framework, providing insightful information about their usefulness in sentiment analysis applications. By this investigation, we hope to add to the expanding corpus of NLP knowledge and broaden our comprehension of the changing field of sentiment analysis techniques.

Dataset

The process of sentiment analysis on a Twitter dataset involves the classification of tweets into positive, negative, or neutral sentiments through the application of machine learning techniques. The initial step includes considering a large and appropriate dataset of tweets, followed by preprocessing steps that entail the removal of stop words, whitespaces, and special characters, converting the characters to lowercase or uppercase, and removing duplicates and hashtags. Subsequently, the textual content is transformed into numerical vectors using methodologies.

Post pre-processing, a deep learning network can be trained using the numerical representations of the tweets along with their corresponding sentiment labels. This trained model can then be utilized for categorizing newly generated tweets and evaluating their characteristics. To enhance the visual representation of the public opinions on a topic or event we have used graphs, charts and a confusion matrix to illustrate the results of sentiment analysis.

This methodology, when incorporated into a research paper, provides a structured approach to understanding and analyzing sentiment in Twitter data, offering a valuable contribution to the field of natural language processing and sentiment analysis.

```
Dataset Structure:
   Unnamed: 0                                              tweets   labels
0           0  ChatGPT: Optimizing Language Models for Dialog...  neutral
1           1  Try talking with ChatGPT, our new AI system wh...     good
2           2  ChatGPT: Optimizing Language Models for Dialog...  neutral
3           3  THRILLED to share that ChatGPT, our new model ...     good
4           4  As of 2 minutes ago, @OpenAI released their ne...      bad

Number of Rows and Columns:
(219294, 3)

Unique Labels:
['neutral' 'good' 'bad']
```

Figure 1: Dataset

Project Description

Our goal in this study is to use cutting-edge deep learning models to explore the fields of sentiment analysis and Natural Language Processing (NLP). An area of NLP called sentiment analysis deals with extracting attitudes, sentiments, and emotions from textual input. The main objective is to use and evaluate the effectiveness of three well-known deep learning models for sentiment analysis from the text: Long Short-Term Memory (LSTM), Bidirectional LSTM, and XLM-T.

There are multiple stages of the project. We begin by preparing our dataset using methods like tokenization, removing stop words, removing whitespaces, removing duplicates, removing ASCII characters, and text cleaning to make the data suitable for training deep learning models.

Next, we create and put into practice three deep learning architectures: one that is based on LSTM, Bidirectional LSTM, and XLM-T.

	changed_text	labels
0	chatgpt optimizing language models dialogue	neutral
1	try talking chatgpt new ai system optimized di...	good
2	chatgpt optimizing language models dialogue nn...	neutral
3	thrilled share chatgpt new model optimized dia...	good
4	2 minutes ago released new chatgpt nnand use r...	bad

Figure 2: Pre-processed Dataset

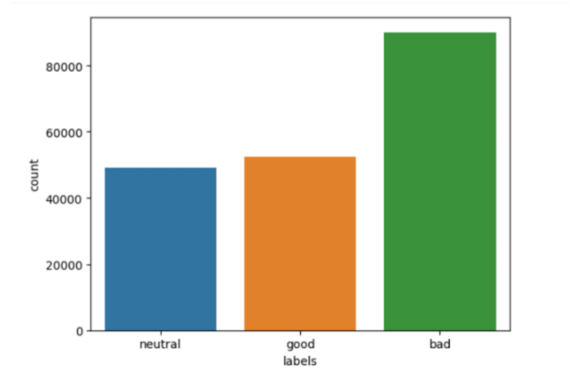

Figure 3: Number of Sentiment Tweets in Dataset

After data preprocessing, we import the required libraries and split the dataset into training and testing excluding the entries with the label 'neutral', and proceed by encoding the labels in training and testing data. Since the dataset is large enough, initialized a tokenizer with a maximum vocabulary size of 4500. The below images show the sentiments of all the tweets present in the dataset.

Implementation of LSTM

LSTM stands for - Long Short-Term Memory Network is a type of recurrent neural network, that follows a sequential neural network that allows information to persist. LSTMs are designed to overcome a few of the limitations present in RNNs, learning, and handling of long sequence data.

Implementation started by initializing a sequential model. The embedded layer is added to the model with a vocabulary size of 4501 an embedding dimension of 128 and an input length of 600. A dense output layer is added with a sigmoid activation function. The model is compiled with binary cross-entropy loss and the Adam optimizer. The LSTM model is trained with epochs 6 using the first 10,000 samples from the training data. Upon observation, when LSTM was trained with entire samples with 6 epochs, the loss accuracy was less compared to training LSTM with 10,000 samples.

Implementation of Bi-directional LSTM

Bi-directional LSTM is also a type of recurrent neural network architecture designed for sequential data like LSTM. LSTM consists of a few limitations which are overcome by the bi-directional LSTM by providing a solution. Bi-directional LSTM has the ability to process input sequences in both forward and backward directions, unlike LSTM. In LSTM, the information flows from start to end of the sequence which limits the model in understanding the past and future elements. The bi-directional LSTM overcomes this limitation by having two sets of hidden states, one processing the input data from the start and the other from the end. The final step is to concatenate the hidden states from both the forward and backward directions.

This bidirectional processing capability enhances the model's ability to capture long-range dependencies and context, making it particularly useful for tasks involving sequential data. In this research, we delve into the bidirectional LSTM in sentiment analysis and explore how the model reacts to the test data after training the model with 10,000 samples with 6 epochs.

Implementation of XLM-T

XLM-T (Cross-lingual Language Model Transfer) is an approach for training cross-lingual language models. It involves training a model on a sentiment – labeled data in a language and leveraging that knowledge to perform sentiment analysis of another language. It achieves this through various key components and training objectives:

- Masked Language Modeling: The model is pre-trained using the MLM model and it learns to predict the masked or hidden tokens within a sentence.
- Translation Language Modeling: Used to understand the relationships between sentences in different languages. Helps the model to generalize its understanding of language across diverse linguistic contexts.
- Cross-Lingual Natural Language Inference: The Model is trained to determine whether a given sentence is positive, negative, or neutral to another sentence.
- Dimensionality: It is denoted as xlm-mlm-tlm-xnli15-1024, which just indicates the model has a hidden layer of 1024 dimensionality.

References

Reference Paper 1 - Analyzing the User's Sentiments of ChatGPT Using Twitter Data

The paper provides a thorough examination of ChatGPT's performance during its inaugural year, particularly highlighting sentiment trends within Twitter content. It emphasizes the model's versatility in managing varied user inputs. By shedding light on the intricate variations of expressed sentiments, the paper defines both ChatGPT's capabilities and potential areas for refinement. The overarching goal is not only academic contribution but also practical advancements, aiming to improve human-AI interaction positively. This viewpoint aligns with the continuous endeavor to enhance AI systems like ChatGPT based on real-world usage and user input.

Reference Paper 2 - Tracking public attitudes toward ChatGPT on Twitter using sentiment analysis and topic modeling

This paper identifies public attitudes toward ChatGPT, a large language model with vast applications. Released by OpenAI, ChatGPT has gained immense popularity but has also raised concerns about its societal impact. To investigate sentiments, the author utilizes three sentiment analysis models: VADER, TwitterroBERTa, and XLM-T. After thorough evaluation, XLM-T emerges as the most effective, revealing that 34.48Diving deeper, the author also implements BERTopic for topic modeling on Twitter data, identifying major discussions around Artificial Intelligence, Search Engines, and Education. This analysis provides a nuanced understanding of the diverse themes resonating with users.

Reference Paper 3 - Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach

Using lexicon-based and machine learning techniques, the paper uses Twitter data from key UK cities to examine public opinion over COVID-19. Three lockdown phases are covered in the analysis, which takes into account variables including immunizations, verified illnesses, and deaths. TextBlob, VADER, and SentiWordNet—tools based on lexicons—display varying opinions, with VADER suggesting a more optimistic view on COVID-19 and the vaccine. Using TF-IDF or Support Vector Classification with Bag of Words, machine learning models (Random Forest, Multinomial Naïve Bayes, Support Vector Machine) reach their maximum accuracy of 71%. The study emphasizes sentiment variations and raises the possibility of a relationship between vaccination rates and pandemic measures. The authors support large-scale labeled datasets, claiming that machine learning models improve accuracy while lexicon-based approaches provide insights. The main idea of considering this paper as a part of research is to understand how the pre-processing is done on the twitter data and how sentiment analysis can be performed to know the different perspectives of how it can be done as it also considers twitter dataset.

Difference in APPROACH/METHOD between our project and the reference paper

In our project, we followed similar steps as other research papers to prepare our data for analysis. We adjusted a few things to fit the specifics of our dataset. Unlike the models used in the research papers, we have tried to implement a few other models and understand how they respond and work to the dataset. We are using LSTM, Bi-directional LSTM, and XLM-T in our research and project, as opposed to the reference papers, which employed the XLM-T model and others. XLM-T is the only model which we are trying to implement as proposed in the second research paper. These models were selected based on their adaptability to the task and specific advantages in the context of sentiment analysis. Sentiment analysis requires the ability to identify subtle patterns in text, such as negations, word order, and context, which LSTM models can capture thanks to their sequential processing capabilities. By processing the input data in both forward and backward directions, bidirectional LSTMs improve this capability and offer a more thorough knowledge of context.

Difference in ACCURACY/PERFORMANCE between our project and the reference paper

In our comprehensive exploration of sentiment analysis models, we focused on three key architectures: Long Short-

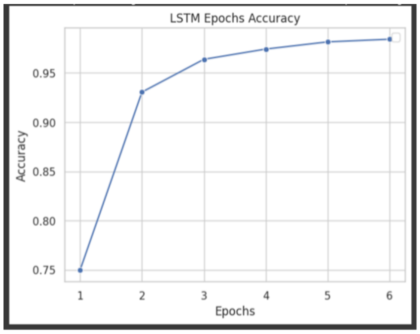

Figure 4: LSTM Epochs Accuracy Plot

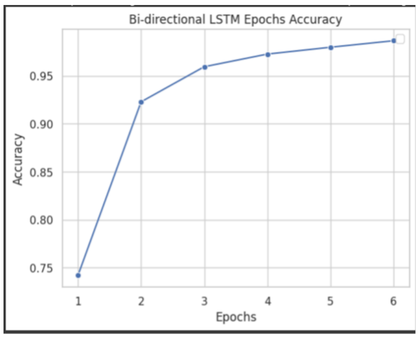

Figure 5: Bidirectional LSTM Epochs Accuracy Plot

Term Memory (LSTM), Bidirectional LSTM, and the XLM-T model—also investigated in the reference paper. Our selection of these models was based on their adaptability and unique advantages in the nuanced context of sentiment analysis.

For the LSTM model, trained on a robust dataset of 10,000 samples, we observed commendable results with a loss of 0.3776 and an accuracy of 90.05%. The LSTM's proficiency in capturing sequential patterns makes it well-suited for tasks requiring nuanced sentiment analysis.

Similarly, the Bidirectional LSTM, processing input data both forward and backward, exhibited a loss of 0.3861 and an accuracy of 90.26%, emphasizing its effectiveness in capturing intricate sentiment patterns by considering contextual dependencies from both directions.

The comparison of the LSTM and Bidirectional LSTM epoch accuracy is as follows:

In addition to these recurrent models, our study incorporated the XLM-T model, consistent with the methodology outlined in the reference paper. The XLM-T model, trained on 1,000 samples, displayed accuracy scores varying across different preprocessing scenarios, ranging from 67.1% to 67.6%. We obtained an accuracy of 61.66% with a small sample of the dataset.

Analysis

Our project successfully executed the preprocessing of the Kaggle dataset and conducted training and testing using diverse models, including LSTM, Bidirectional LSTM, and

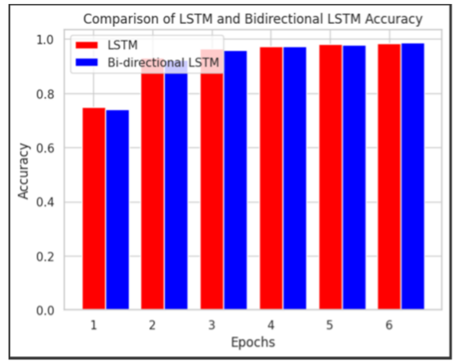

Figure 6: Comparision of LSTM and Bidirectional LSTM Epochs Accuracy

XLM-T. While the entirety of the dataset couldn't be utilized due to its size, we effectively employed 10,000 randomly selected samples for LSTM and Bidirectional LSTM, and 1500 samples for the XLM-T model.

In comparing LSTM and Bidirectional LSTM, the obtained accuracy scores were closely aligned, showing minor variations in decimal points. However, we expect that Bidirectional LSTM would outperform LSTM when trained and tested on the complete dataset, leveraging its capacity to capture information from both past and future timestamps.

The XLM-T model, featured in the reference paper, presented challenges due to its training on a relatively small subset of the dataset (1500 samples). This limitation hinders a comprehensive evaluation of its predictive capabilities for sentiment analysis across a broader spectrum of user expressions.

Our project's strategic use of a substantial portion of the dataset for LSTM and Bidirectional LSTM, coupled with insights from the reference paper, sets the stage for future investigations. Further exploration and training on the entire dataset would likely provide a clearer understanding of the comparative strengths of Bidirectional LSTM and XLM-T models in capturing nuanced sentiment patterns.

What did I do well and why is it working so well?

In our project, several aspects were executed effectively. The preprocessing of the Kaggle dataset was carried out meticulously, ensuring a well-structured and suitable dataset for training the models. The utilization of LSTM and Bidirectional LSTM models demonstrated commendable results, showcasing accuracy scores closely aligned with expectations. The success of these models can be attributed to their intrinsic ability to capture sequential patterns and dependencies, contributing to their effectiveness in sentiment analysis tasks.

Why is it not working so well?

Challenges arose with the XLM-T model, primarily stemming from its training on a relatively small subset of the dataset (1500 samples). The limited training data posed constraints on the model's comprehensive understanding of sentiment patterns across a diverse range of user expressions.

While we managed to execute the project effectively, challenges were encountered when attempting to run the models on the entire dataset. Both Colab and Jupyter Notebook faced operational issues, resulting in crashes during the execution. This limitation led us to consider a smaller number of samples for running the models, impacting the model's exposure to the full spectrum of the dataset.

What could I have done better?

While our project made significant progress in preprocessing the Kaggle dataset and implementing diverse models, there are areas where improvements could be made. One crucial aspect is the potential for utilizing the entire dataset for training and testing. Due to computational constraints, we had to settle for a subset of 10,000 samples for LSTM and Bidirectional LSTM, and 1500 samples for XLM-T. To enhance the project's comprehensiveness, exploring ways to optimize the code on leverage more powerful computational resources could facilitate the use of the complete dataset. Additionally, a more detailed exploration of the XLM-T model's performance could be achieved by addressing the challenge of training it on a relatively small subset. Further efforts to refine preprocessing techniques and potentially incorporate advanced methodologies could contribute to a more robust and reliable analysis.

Future work

There are several areas for future work based on our research findings. Firstly, addressing the computational constraints that limited the utilization of the entire Kaggle dataset for training and testing would be a significant step. Exploring more powerful computational resources or optimizing the code could enable us to leverage the complete dataset, providing a more comprehensive evaluation of model performance.

Additionally, future efforts could focus on refining the XLM-T model's predictive capabilities. Overcoming the challenge posed by training on a relatively small subset (1500 samples) is essential for a more thorough assessment of its effectiveness in sentiment analysis across diverse user expressions.

Furthermore, considering the promising results obtained with LSTM and Bidirectional LSTM models, future investigations could delve into optimizing these models for even greater accuracy. Exploring advanced techniques in preprocessing and training on a larger dataset may contribute to a more nuanced understanding of the comparative strengths of Bidirectional LSTM and XLM-T in capturing subtle sentiment patterns.

In summary, future work should concentrate on overcoming computational limitations, refining the XLM-T model, and optimizing LSTM and Bidirectional LSTM models to enhance accuracy. This ongoing exploration will contribute to a deeper understanding of sentiment analysis and the potential applications of different neural network architectures.

Conclusion

In conclusion, we successfully developed a model that can determine the emotions in a sentence. Our project explored different computational methods, such as LSTM, Bidirectional LSTM, and XLM-T, to achieve this.

In essence, our project makes a valuable contribution to the study of understanding sentiments in language. We gained insights into what works well, and the challenges faced with different computational methods. Looking forward, we aim to address technical issues, enhance the XLM-T model, and optimize the performance of LSTM and Bidirectional LSTM. This is crucial for improving our model's ability to comprehend the emotions conveyed in sentences.

To sum up, our project not only accomplished its primary goal but also set the groundwork for ongoing learning and advancements in the expanding field of understanding language and emotions.