

Comment Sentiment Analysis Using Bidirectional Encoder Representations from Transformers

Harshit Mittal^{1*}, Dr. Neeraj Garg²

¹Student, Computer Science Department, Maharaja Agrasen Institute of Technology, Delhi, India

²HOD, AI/ML Department, Maharaja Agrasen Institute of Technology, Delhi, India.

*Corresponding Author

Harshit Mittal

Student, Computer Science
Department, Maharaja
Agrasen Institute of
Technology, Delhi, India.

Article History

Received: 11.12.2023

Accepted: 05.01.2024

Published: 04.02.2024

Abstract: This research paper unravels the power of Bidirectional Encoder Representations from Transformers (BERT), a cutting-edge language representation model, in the realm of comment sentiment analysis. By focusing on two main aspects the working principles of BERT and the methodology of comment sentiment analysis - the paper aims to captivate readers with an engaging exploration of this exciting field. In this paper, we delve into the remarkable workings of BERT, showcasing its exceptional ability to grasp context, semantics, and sentiment in natural language text. Through a comprehensive examination of its underlying architecture, intricate pre-training process, and ingenious fine-tuning techniques, readers will gain a profound understanding of how BERT transforms raw language into rich, contextualized representations. Building upon this foundation, the paper unveils the methodology employed in our comment sentiment analysis project. Here, we take readers on an enthralling journey through the data collection and pre-processing phases, carefully selecting appropriate evaluation metrics, and ingeniously designing a sentiment classification framework. Along the way, we address the challenges of word sense disambiguation and tackle the nuances of neutral sentiments. By venturing into the uncharted territory of comment sentiment analysis using BERT, this research paper becomes more than just an invaluable resource for practitioners and researchers. It serves as a seductive invitation to unlock the hidden treasures within user opinions and sentiments, enabling us to leverage this invaluable wealth of information across various domains. With insightful findings and actionable recommendations, this paper fuels the pursuit of advanced sentiment analysis techniques, propelling the field towards enhanced comprehension and optimal utilization of user sentiments in the ever-evolving digital landscape.

Keywords: Bidirectional Encoder Representations from Transformers (BERT), Comment Sentiment Analysis, Natural Language Processing (NLP), Data Pre-processing.

I. Introduction

In the era of digital communication, the internet has become an invaluable repository of user opinions and sentiments. From product reviews to social media comments, these user-generated texts hold a wealth of information that can shape businesses, influence public perception, and aid decisionmaking.

Extracting meaningful insights from this vast sea of text data has become a vital endeavor, leading researchers and practitioners to explore cutting-edge techniques in sentiment analysis. One such technique that has captured the attention of the natural language processing community is BERT – Bidirectional Encoder Representations from Transformers[14]. BERT represents a groundbreaking approach to language representation, revolutionizing the way we understand and process textual information. It has proven to be a formidable force in various language understanding tasks, including sentiment analysis. This research paper embarks on a fascinating journey to explore the synergy between BERT and comment sentiment analysis[5,6]. With a twofold focus on understanding the inner workings of the BERT model and the methodology of comment sentiment analysis,

we aim to provide readers with an enlightening and enriching experience.

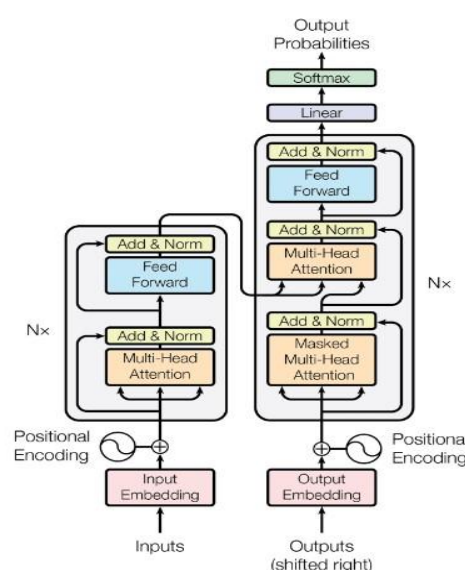


Fig. 1. Basic Transformer Model (Courtesy: Google).

Starting the paper with an introduction, we moved to the second part where we delve deep into the inception and development of BERT[1-4], uncovering its architectural intricacies, its pre-training process, and its remarkable ability to capture the complex nuances of human language. Understanding the inner workings of BERT is crucial for leveraging its power effectively and extracting meaningful sentiment analysis insights.

Building upon this foundation, the third part of the paper presents the methodology employed in our comment sentiment analysis project. We guide readers through the meticulous process of data collection and pre-processing to ensure the creation of a clean and reliable dataset. Next, we explore the selection of appropriate evaluation metrics to assess the performance of sentiment analysis models. Finally, we present our innovative sentiment classification framework, emphasizing techniques to handle word sense disambiguation challenges and effectively discern neutral sentiments.

In the fourth part of the paper, we showcase the success of the model through our findings and results. Overall, through this research, we aspire to ignite a spark of inspiration in the minds of readers, fostering further advancements in sentiment analysis techniques and enhancing our ability to comprehend and harness the power of user opinions in various domains.

II. UNDERSTANDING BERT

A. History and Relation to Other Models

The emergence of BERT in 2018 marked a watershed moment in Natural Language Processing (NLP)[1-6]. But to truly understand its significance, we must first journey back in time and explore the fertile ground that nurtured its birth. This section delves into the historical context and key predecessors that paved the way for BERT's groundbreaking architecture. Prior to BERT, the NLP landscape was dominated by recurrent neural networks (RNNs)[1-4] and convolutional neural networks (CNNs)[7-10]. While these models achieved impressive results in various tasks, they also had inherent limitations that hindered their ability to fully grasp the nuances of language. These limitations fueled the need for a new approach, one that could effectively capture the intricate web of connections within language. Enter the Transformer architecture, the stage upon which BERT would rise to stardom.

In 2017, the paper "Attention Is All You Need" introduced the Transformer architecture, a groundbreaking approach that fundamentally shifted the NLP paradigm. The core of this architecture lay in the self-attention mechanism, a powerful tool that allowed the model to attend to all parts of an input sentence simultaneously, rather than sequentially. Imagine the self-attention mechanism as a skilled translator, instantly grasping the relationships between all the words in a sentence, just like a human reader. This newfound ability to consider long-range dependencies and global context was a game-changer, paving the way for significantly more nuanced and accurate language understanding. Inspired by the Transformer's potential, the Google AI team in 2018 unveiled BERT (Bidirectional Encoder Representations from Transformers). BERT took the Transformer architecture to the next level by introducing several key innovations:

Bidirectional training: Unlike most previous models that processed text unidirectionally (left to right or right to left), BERT trained on both directions simultaneously. This allowed it to

capture even richer contextual information, considering not only the preceding words but also those that followed. Masking mechanisms: To prevent the model from "cheating" by peeking at future words during training, BERT employed clever masking techniques that ensured it learned to rely solely on the available context. Pre-training on massive datasets: BERT was pre-trained on a massive dataset of text and code, allowing it to acquire a vast understanding of language patterns and nuances. This pre-trained model could then be fine-tuned for specific NLP tasks, achieving remarkable performance gains. The combination of these advancements propelled BERT to the forefront of NLP research, achieving state-of-the-art results in a wide range of tasks, including sentiment analysis, question answering, and text summarization.

BERT

Encoder

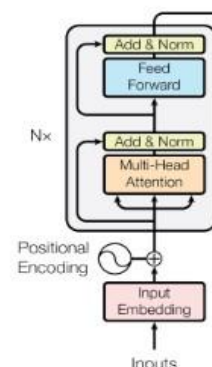


Fig. 2. Bert Basic Block or Encoder Block (Courtesy: Google).

B. Architectural and Algorithmic Insights

BERT's groundbreaking performance hinges on its unique architecture, a masterful blend of the Transformer encoder and pre-training methodologies. Understanding the intricate workings of this engine is key to appreciating its significance and impact on the NLP landscape.

At the heart of BERT lies the Transformer encoder, a multilayered stack of identical "Transformer blocks." Each block employs the powerful self-attention mechanism, allowing the model to analyze the relationships between every word in a sentence simultaneously, unlike RNNs' sequential processing. Imagine the Transformer encoder as a bustling marketplace where words can freely interact and exchange information. The self-attention mechanism acts as a skilled mediator, ensuring each word understands the context and contribution of every other word in the sentence. Each Transformer block within the encoder comprises several sub-processes that work in concert:

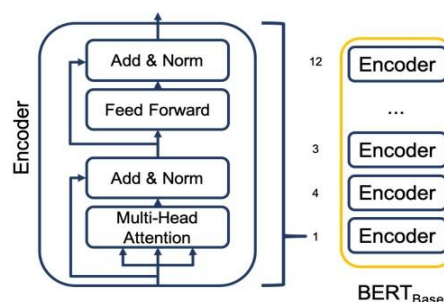


Fig. 3. BERT Structure (Courtesy: Google).

Word Embeddings: The first step in BERT's journey involves transforming words from their raw textual form into numerical representations called word embeddings. These highdimensional vectors capture the semantic meaning of words, allowing the model to quantify and manipulate relationships between them mathematically. Imagine a vast library where each book represents a unique word, and its position on the shelves reflects its meaning. Word embeddings create this conceptual library, mapping words into a structured vector space.

Positional Encodings: Language is inherently sequential, and the order of words matters for understanding meaning. To ensure BERT grasps this, positional encodings are added to the word embeddings. These encodings inject information about each word's position within the sentence, providing valuable contextual cues. Think of these encodings as timestamps added to each book in the library, indicating when it was placed on the shelf. This temporal information helps BERT reconstruct the original sentence flow and context.

Multi-Head Attention: The core algorithm that sets BERT apart is multi-head attention. This mechanism allows the model to attend to different parts of the sentence simultaneously, capturing long-range dependencies and complex relationships between words. Think of multi-head attention as a team of expert librarians working together, each focusing on a specific aspect of the library's collection. They can crossreference information, uncover hidden connections, and build a comprehensive understanding of the semantic landscape. Here's how multi-head attention operates within each Transformer block:

Query, Key, and Value Vectors: Each word's embedding is split into three vectors: query, key, and value. **Scoring:** The query vector of each word is compared to the key vectors of all other words using dot-product similarity scores. These scores reflect the degree of relevance between words. **Softmax:** The scores are normalized using the softmax function, ensuring they sum to 1, creating a probability distribution of attention weights. **Weighted Sum:** The value vectors are multiplied by their corresponding attention weights and summed together, generating a context-aware representation for each word. **Multiple Heads:** Multiple attention heads perform this process in parallel, each attending to different aspects of the sentence, capturing a richer understanding.

Feed-forward Networks: After each attention layer, BERT employs feed-forward networks to further refine the encoded information. These fully connected layers add non-linearity and deeper contextual understanding to the representations. Imagine these networks as wise scholars who analyze the knowledge gathered by the librarians, extracting deeper insights and synthesizing a more comprehensive understanding of the text. Through multiple iterations of these processes, the Transformer encoder builds a rich understanding of the entire sentence, capturing intricate contextual nuances and relationships between words.

Pre-training: After each attention layer, BERT employs feed-forward networks to further refine the encoded information. These fully connected layers add non-linearity and deeper contextual understanding to the representations. Imagine these networks as wise scholars who analyze the knowledge gathered by the librarians, extracting deeper insights and synthesizing a more comprehensive understanding of the text.

C. Strengths and Limitations

BERT's emergence has undoubtedly transformed the NLP landscape, but like any powerful tool, it comes with its own set of strengths and limitations. Understanding both sides provides a nuanced perspective on its impact and future potential.

Strengths:

Contextual understanding: BERT's bidirectional processing and self-attention mechanism allow it to grasp the intricate relationships between words within a sentence and across phrases. This contextual awareness leads to significantly more accurate interpretations compared to traditional models. **Versatility:** BERT's pre-trained model excels in a wide range of NLP tasks, including sentiment analysis, question answering, text summarization, and more. This makes it a valuable tool for researchers and developers across various fields. **Transfer learning:** BERT's pre-trained knowledge can be fine-tuned for specific tasks, reducing training time and resources significantly. This democratizes access to advanced NLP capabilities for diverse applications. **State-of-the-art performance:** BERT consistently achieves top results on benchmark NLP tasks, pushing the boundaries of what's possible in language understanding. This continuous advancement inspires further innovation and refinement in the field.

Limitations:

Computational complexity: Training and running BERT models require significant computational resources, making them less accessible for low-powered devices and environments. This limits its potential for real-world deployment in certain scenarios. **Black-box nature:** The inner workings of BERT, particularly the multi-head attention mechanism, can be difficult to interpret and understand. This poses challenges in debugging and ensuring fairness and trustworthiness in its outputs. **Data bias:** BERT's performance can be influenced by the biases present in the training data it's exposed to.

This raises concerns about potential perpetuation of discrimination and unfairness in its applications. **Explainability and interpretability:** While BERT excels at generating outputs, providing clear explanations for its reasoning and decisionmaking process remains a challenge. This hinders transparency and trust in its applications for critical tasks.

Despite its limitations, BERT's strengths undeniably outweigh its drawbacks. It has laid a solid foundation for further advancements in NLP, and ongoing research efforts are actively addressing its limitations. Continued work on explainability, bias mitigation, and efficiency improvements will unlock its full potential for diverse real-world applications.

III. METHODOLOGY

A. Data Gathering and Understanding

For this paper, we gathered comments and posts with a good number of engaging. We collected around 15000 comments and categorized them manually into 3 classes negative, positive, and neutral. The word count of each comment lies between the average of 20 to 25 words which is normalized to increase the quality of the model.

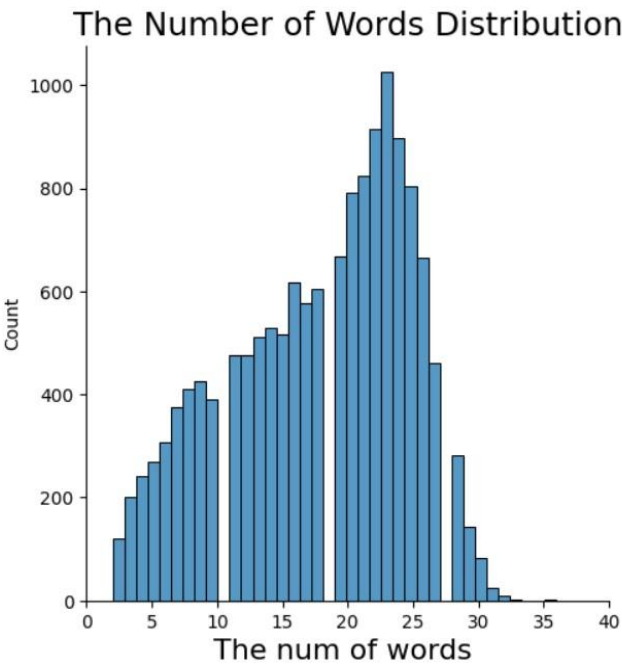


Fig. 4. Word count comparison.

Initially, Every class is distributed in an unequal manner where Negative contains 9000 words, and positive and neutral contain 3000 words each. So, to bring them in an equal ratio and prevent the over-pressure on the CPU, the total number of words brought to 5000 comments, which drastically increased the speed and quality of training.

B. Data Pre-processing

As we went through the different comments we removed some extra characters which are useless and confusing for our model, for example: "I flying @VirginAmerica", and "@AsisAirlines Thanks!". We removed the emojis, hashtags, mentions with '@', and URLs from these comments. To further make our data more understandable by model, we turned the entire comment into lowercase and removed extraneous spaces.

text	cleaned_text
@VirginAmerica What @dhepburn said.	what said
@VirginAmerica plus you've added commercials t...	plus youve added commercials to the experience...
@VirginAmerica I didn't today... Must mean I n...	i didnt today must mean i need to take another...
@VirginAmerica it's really aggressive to blast...	its really aggressive to blast obnoxious enter...
@VirginAmerica and it's a really big bad thing...	and its a really big bad thing about it
@VirginAmerica seriously would pay \$30 a fligh...	seriously would pay 30 a flight for seats that...
@VirginAmerica yes, nearly every time I fly VX...	yes nearly every time i fly vx this ear worm w...
@VirginAmerica Really missed a prime opportuni...	really missed a prime opportunity for men with...
@virginamerica Well, I didn't...but NOW I DO! :-D	well i didntbut now i do d
@VirginAmerica it was amazing, and arrived an ...	it was amazing and arrived an hour early youre...
@VirginAmerica did you know that suicide is th...	did you know that suicide is the second leadin...
@VirginAmerica I <t3 pretty graphics. so muc...	i lt3 pretty graphics so much better than mini...

Fig. 5. Cleansed Text.

The semantics class in the dataset has to be turned from strings 'Negative', 'Positive', and 'Neutral' to numerical values '2', '1', and '0' respectively, to further simplify the dataset for the model to be more efficient.

airline_sentiment	NumericValue
neutral	0
positive	1
neutral	0
negative	2
negative	2

Fig. 6. Classes in form of integers.

C. Model Selection

The success of comment semantics analysis heavily relies on the choice of an appropriate model. Choosing the right model for sentiment analysis can be tough, but DistilBERT stands out as a strong contender for analyzing the often nuanced and dynamic sentiments of Twitter comments. Here's why:

Efficiency and Speed: DistilBERT is a smaller and faster version of the original BERT model, making it perfect for resource-constrained environments like sentiment analysis on large datasets of tweets. Its reduced size allows for quicker training and analysis, saving you valuable time and computational resources.

Contextual Understanding: DistilBERT captures the intricate relationships between words within a sentence and across phrases, crucial for understanding the often-ambiguous sentiment in informal Twitter language. It goes beyond surfacelevel word analysis and grasps the context and nuances of how words are used.

Handling Sarcasm and Irony: Twitter is notorious for its sarcasm and irony, which can trip up traditional sentiment analysis models. DistilBERT's pre-trained knowledge and powerful attention mechanism help it better identify these linguistic subtleties and interpret the intended sentiment correctly.

Transfer Learning and Fine-tuning: DistilBERT's pretrained model can be easily fine-tuned for specific tasks like sentiment analysis on Twitter data. This saves you the time and effort of training a model from scratch and leverages its prior knowledge of language patterns and nuances.

Multilingual Capabilities: If your Twitter data encompasses multiple languages, DistilBERT can be a valuable asset. Its multilingual models can handle diverse languages with accuracy, ensuring you capture the sentiment across different linguistic landscapes.

Open-Source and Accessible: DistilBERT is freely available as an open-source model, making it accessible to researchers and developers without licensing restrictions. This allows for greater collaboration and easier integration into your existing projects.

D. Data Tokenization

Tokenization is a pivotal step in preparing textual data for input into the BERT model. Unlike traditional models that treat each word as a separate entity, BERT operates on subword tokens,

allowing it to capture fine-grained semantic nuances. This process involves breaking down input sentences into smaller units, often words or subwords, to create a sequence of tokens. For this project, we used pre-trained tokenizer "distilbert-base-uncased" which employs a WordPiece tokenizer, words are further split into subword tokens. This approach enables the model to handle out-of-vocabulary words and understand the context of subword variations.

During tokenization, careful consideration is given to preserving the structure and meaning of comments. Special tokens, such as [CLS] (classification) and [SEP] (separator), are added to mark the beginning and end of each comment. Additionally, tokens may be assigned unique identifiers, facilitating the reconstruction of the original comment during post-processing.

Given the fixed-length input requirements of BERT, comments longer than the model's maximum token limit need to be truncated. Once tokenized and possibly truncated, the comment is encoded into numerical representations that the BERT model can process. Each token is mapped to its corresponding index in the BERT vocabulary, creating a numerical sequence. Additionally, segment embeddings and position embeddings are added to the tokenized input. Segment embeddings help the model distinguish between different segments within a comment, while position embeddings provide information about the position of each token in the sequence.

E. Model Training

The next step is fine-tuning the model for the specific task at hand. Given the nature of comment semantics analysis, we used a dataset annotated with labeled comments representing different semantic categories. The model is trained to predict these labels, adjusting its parameters to best capture the underlying semantic structure of the comments as shown below:

total number of epochs=2, per device train batch size=8, per device eval batch size=16, warmup steps=500, weight decay=0.01, logging dir='./logs', logging steps=10, eval steps=10

Fine-tuning involves adapting the pre-trained BERT model to the intricacies of the comment semantics task. We experimented with different hyperparameters, batch sizes, and learning rates to optimize the model's performance. The training was conducted on a suitable hardware setup, ensuring efficient processing and convergence of the model.

For training model was divided into two parts, 80 percent of the data is provided for training data and the rest 20 percent for testing.

IV. RESULTS AND DECLARATIONS

The application of the trained model to the test dataset yielded promising results. Comparison Metrics were computed to measure the model's performance across various semantic categories. The result from metrics was outstanding, it showed that our model was able to predict every true label of comment giving the model 100 percent accuracy.

```
array([[209, 0, 0],
       [ 0, 183, 0],
       [ 0, 0, 608]])
```

Fig. 7. Comparison Matrix.

Beyond numerical metrics, qualitative analysis of the model's predictions on specific comments offers valuable insights. Examining instances where the model excelled or struggled can shed light on its interpretability and help identify patterns or challenges in comment semantics analysis. Moreover, analyzing misclassifications can guide future improvements in the model.

V. CONCLUSION

In this paper, we embarked on a comprehensive exploration of comment semantics analysis utilizing the powerful BERT model. Our journey began with an in-depth introduction to BERT, emphasizing its bidirectional attention mechanism and pre-trained contextual understanding of language. Leveraging BERT's capabilities, we designed a robust methodology for comment semantics analysis and presented each crucial step with meticulous detail. The choice of BERT as our model of choice was grounded in its proven effectiveness across a myriad of natural language processing tasks. The model demonstrated its prowess in capturing nuanced contextual relationships within comments, a crucial aspect of semantic analysis in user-generated content. We meticulously addressed the intricacies of data collection and pre-processing, ensuring the prepared dataset was aptly suited for the task. Tokenization, truncation, and encoding processes were discussed in detail, shedding light on the transformation of raw textual data into a format compatible with BERT's input requirements. The methodology's efficacy was then put to the test through model training and evaluation. We fine-tuned BERT on a labeled dataset, showcasing its prowess in discerning diverse semantic categories within comments. The results, presented quantitatively through comparison metrics, provided a comprehensive view of the model's 100 percent accuracy. This paper serves as a testament to the effectiveness of BERT in unraveling the complex tapestry of comment semantics. As we navigate the ever-evolving landscape of natural language processing, the lessons learned from this study pave the way for future advancements and applications in understanding the rich semantics embedded in user comments across diverse digital platforms.

VI. MESSAGE FROM AUTHORS

The corresponding author "Harshit Mittal" would like to thank "Dr. Neeraj Garg" and Maharaja Agrasen Institute of Technology Delhi for their constant support and supervision in helping him write this paper.

REFERENCES

1. Lin, T., Wang, Y., Liu, X., & Qiu, X. (2022). A survey of transformers. *AI open*, 3, 111-132.
2. Fournier, Q., Caron, G. M., & Aloise, D. (2023). A practical survey on faster and lighter transformers. *ACM Computing Surveys*, 55(14s), 1-40.
3. Koroteev, M. V. (2021). BERT: a review of applications in natural language processing and understanding. *arXiv preprint arXiv:2103.11943*.
4. Lin, J., Nogueira, R., & Yates, A. (2022). *Pretrained transformers for text ranking: Bert and beyond*. Springer Nature.
5. Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. (2023). Sentiment analysis and opinion mining on educational data: A survey. *Natural Language Processing Journal*, 2, 100003.

6. Mehra, P. (2023). Unexpected surprise: Emotion analysis and aspect based sentiment analysis (ABSA) of user generated comments to study behavioral intentions of tourists. *Tourism Management Perspectives*, 45, 101063.
7. Mittal, H. (2023). Evaluating The Performance of Feature Extraction Techniques Using Classification Techniques. *Computer Science & Information Technology (CS & IT)*, 13(14), 19-20.
8. Mittal, H., & Garg, N. (2022, February). Recognizing/Detecting Human Faces in Images: Survey. In *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*.
9. Mittal, H., & Garg, N. (2023). A Survey On Automatic Vehicles Using Computer Vision. Available at SSRN 4387778.
10. Mittal, H. (2023). Kidney CT Image Analysis Using CNN. *Computer Science & InformationTechnology (CS & IT) ISSN*, 2231-5403.