AMERICAN Journal of Engineering,
Mechanics and Architecture
Volume 3, Issue 4, 2025 ISSN (E): 2993-2637

Data Analysis Based on the Bert Model

Eshankulov Hamza Ilhomovich
Professor of the Department of Applied Mathematics and Programming Technologies at BSU
x.i.eshonkulov@buxdu.uz

Soliyeva Dilsora Alisher qizi
Master's Degree Holder at the Department of Applied Mathematics and Programming Technologies at BSU
soliyeva.dilsora.2020@gmail.com

Abstract: This article provides an in-depth analysis of the capabilities of the BERT (Bidirectional Encoder Representations from Transformers) model, which holds a significant place in the field of Natural Language Processing (NLP), for performing deep, contextual, and semantic data analysis. The internal architecture and working mechanisms of the BERT model are explored, along with advanced approaches such as DistilBERT and Topic BERT-BiGRU, focusing on methodologies for accurately classifying both complex and short texts[5]. The paper also presents a comparative analysis with traditional models, clearly outlining the advantages of modern approaches. [1]

Main Results: According to the research findings, approaches based on the BERT model demonstrate high accuracy in contextual and semantic text analysis. Specifically, the Topic BERT-BiGRU model stood out with an F1-score of 86.91% on short texts. DistilBERT outperformed the classic BERT model by 0.8–1.2% in accuracy on smaller datasets. Transformer-based classification systems, especially when combined with the Random Forest classifier, achieved over 90% accuracy. These approaches confirm the high efficiency of the BERT model in real-world applications.

Keywords: BERT, Transformer, BiGRU, topic modeling, NLP, semantic analysis, text classification.

I. INTRODUCTION

In the era of modern information technologies, the volume of textual data obtained from the internet, social networks, and online platforms is rapidly increasing. The need to automatically analyze, classify, and draw logical conclusions from this data has become a strategic direction in the field of Natural Language Processing (NLP). Traditional approaches, including Bag-of-Words (BoW), TF-IDF, and Word2Vec models, have key limitations—primarily their inability to fully capture the semantic relationships between words and to sufficiently understand context. [6] BERT, on the other hand, was developed precisely to address these issues, offering an advanced approach that deeply understands texts through bidirectional encoding, self-attention, and pre-training mechanisms. [5]

In today's digital world, the volume of textual content on the internet is growing at an exponential rate. At the same time, users' language-related needs—such as automatic translation, text summarization, chatbots, recommendation systems, and fake news detection—are also rapidly increasing. Traditional methods have proven inadequate when dealing with such complex and multi-layered texts. In this context, transformer models based on deep learning, such as BERT, have gained even greater significance. The BERT model has become a revolutionary solution for understanding contextual semantics, detecting polysemy, and capturing complex relationships between linguistic units. Therefore, this topic is considered highly relevant not only from a scientific perspective but also in practical applications.

The main goal of this study is to thoroughly explore the role of the BERT model in natural language processing, particularly in text classification and analysis, to examine its technical mechanisms, and to scientifically compare the advantages of improved approaches such as DistilBERT and Topic BERT-BiGRU. This article analyzes the performance, capabilities, and limitations of BERT-based models by applying them to different types of text corpora. As a result, a scientific and methodological foundation is established for the practical application of the BERT model.

II BERT ARCHITECTURE AND ITS APPLICATION IN TEXT CLASSIFICATION

2.1. Internal Structure and Working Principle of the BERT Model. The BERT model was introduced by Google AI in 2018. Its main advantage lies in the ability to understand the meaning of each word not only in the context of preceding words but also in the context of subsequent ones. This is made possible through two pre-training strategies: Masked Language Modeling (MLM) and Next Sentence Prediction (NSP). Built on the Transformer architecture, BERT represents words as vectors using complex attention mechanisms, where these vectors encapsulate deep contextual semantic information. [4] BERT models have demonstrated state-of-the-art performance in various NLP tasks, such as question answering, text classification, and named entity recognition.

BERT is a deep neural network-based natural language understanding model that can learn the semantics and structure of language from large volumes of unlabeled text. [5] One of BERT's core innovations is the use of a bidirectional Transformer encoder, which considers both left and right contexts to extract richer linguistic features. The model's pre-training tasks—MLM and NSP—are designed to learn representations at the vocabulary and sentence levels, respectively. This model utilizes only the encoder part of the Transformer architecture and operates by incorporating bidirectional context. (See Figure 1)

Input Layer: When text is input into the model, it is converted into a numerical representation using three types of embeddings:

- **Token Embedding** – a unique vector for each token;
- **Segment Embedding** – used to distinguish between different text segments;
- **Position Embedding** – represents the position of tokens within the text.

These embeddings are combined as follows:

$\mathrm{E} = \mathrm{E}_{\mathrm{token}} + \mathrm{E}_{\mathrm{segment}} + \mathrm{E}_{\mathrm{position}}$ (1)

Transformer Encoder Blocks: The BERT model consists of multiple encoder blocks. Each block includes the following components:

- **Multi-head Self-Attention:** Each token is connected with all other tokens in the input sequence, allowing the model to capture contextual relationships.
- **Feed Forward Neural Network (FFNN):** Independently processes each token through a fully connected neural network.
- **Residual Connections and Layer Normalization:** These components ensure the stability and efficiency of the model during training by preserving gradients and normalizing activations.

Each encoder block can be represented by the following formulas:

$\mathrm{Z} = \mathrm{LayerNorm}(\mathrm{X} + \mathrm{MultiHeadAttention}(\mathrm{X}))$ (2)

$H = LayerNorm(Z + FFNN(Z))$ (3)

Here:

- **X** – Input vector
- **Z** – Result of the self-attention mechanism
- **H** – Output of the encoder block

Output Layer: At the output of the model, the vector corresponding to the [CLS] token is taken as a general representation of the entire input text. This vector is then passed through the following classification layer:

$\boldsymbol{y = softmax(W * h + b)}$ (4)

Here:

- **h** – Hidden vector of the [CLS] token
- **W** – Weight matrix
- **b** – Bias vector
- **y** – Output probabilities

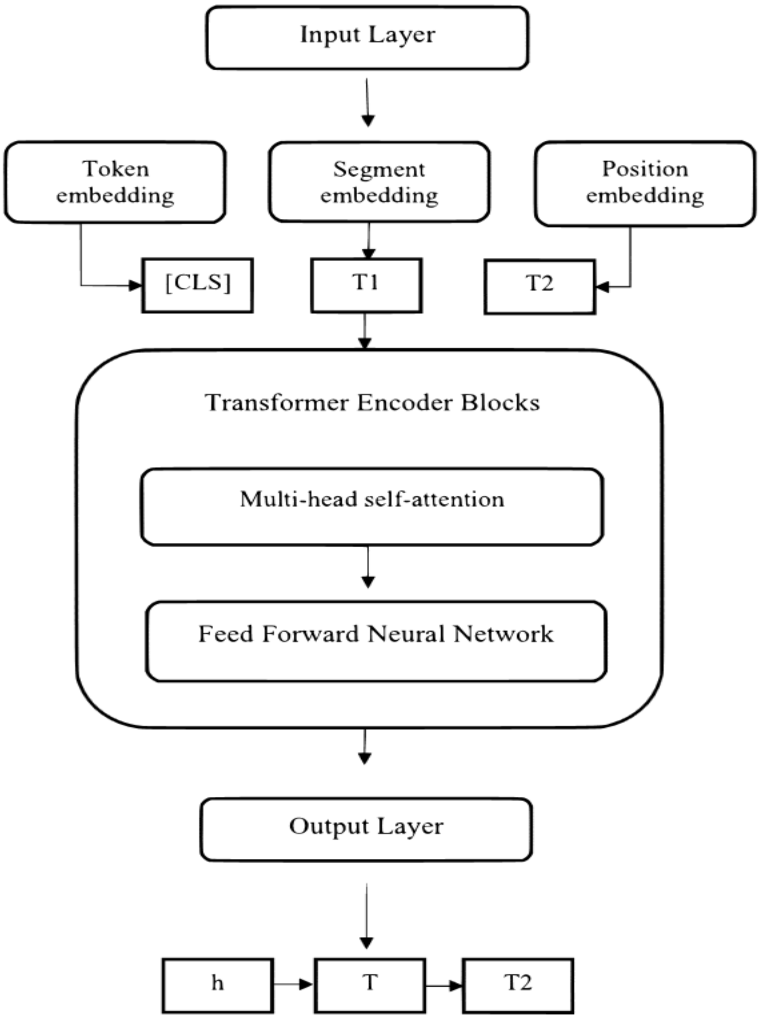

Figure 1.

2.2. Text Classification Based on BERT. The process of text classification using the BERT model consists of two stages: pre-training and fine-tuning. During the pre-training phase, the model learns general language understanding from large corpora of text. In the fine-tuning phase, the model is adapted to a specific task. For classification, the [CLS] token is taken as the main semantic vector, and the class is determined through a softmax function. This approach demonstrates high performance, especially when dealing with texts containing polysemous words and complex sentence structures.

III. ADVANCED MODELS BASED ON BERT AND THEIR PRACTICAL PERFORMANCE

DistilBERT: The DistilBERT model enriches the semantic analysis capabilities of BERT by incorporating rule-based logic based on an additional lexicon. [6] The core idea of the model is that if the ratio of positive-to-negative probabilities in the BERT output is low, it indicates that the model has produced unreliable results. In this case, a sentiment lexicon-based rule is applied to improve the result's accuracy. This approach, especially in small datasets, provides an improvement of 0.8–1.2% over traditional BERT.

Topic BERT-BiGRU: Short texts, particularly those in titles, comments, and tweets, reveal the limitations of the BERT model. This is because the Masked Language Modeling (MLM) strategy can lead to the loss of semantic richness when masking keywords in short texts. To address this, the Topic BERT-BiGRU model was developed, which enhances BERT embeddings with Biterm Topic Model (BTM) topic vectors. [7] BiGRU captures semantic dependencies in both directions of a sequence and extracts deep features. The model's performance is demonstrated by an F1 score of 86.91% on the THUCNews dataset.

IV. MODEL RESULTS AND ANALYTICAL COMPARISON

Intellectual Analysis: The classification system based on a transformer model classified using BERT has achieved high accuracy on large text corpora in real-life scenarios, including the Reuter and Daily Star datasets. Specifically, when combined with the Random Forest classifier, the BERT model's embeddings have shown accuracy exceeding 90%. [2] This demonstrates the practical advantages of the BERT model based on contextual semantics.

№	Model Name	Dataset	Accuracy (%)/F1-score
1	BERT	THUCNews	83,29%
2	BERT+BiGRU	THUCNews	83,95%
3	Topic BERT-BiGRU	THUCNews	86,91%
4	DistilBERT	Sentiment dataset	95,20%
5	BERT	Sentiment dataset	94,38%
6	BERT+Random Forest	Sentiment dataset	90%

The research results show that models developed based on BERT demonstrate varying levels of effectiveness across different types of text. Specifically, the Topic BERT-BiGRU model shows significant superiority in working with short texts, recording an F1-score of 86.91% on the THUCNews headline dataset. This is 3.62% higher than the standard BERT model, demonstrating the effectiveness of combining topic vectors and bidirectional sequences in the context of short texts.

The DistilBERT model, on the other hand, showed high accuracy on small-sized and sentiment-based datasets. By evaluating the reliability of BERT outputs, the model refers to an adaptive lexicon when necessary. This approach provided a high performance with an F1-score of 95.20%.

Additionally, the integration of the BERT model with classical classifiers has demonstrated high performance in classification tasks across broader domains. In systems such as those working with Reuter and Daily Star corpora, accuracy was shown to exceed 90%.

These results indicate that the BERT model, with its fundamental architecture, can perform high-level semantic analysis. However, by enhancing it with specialized architectures such as topic modeling, sequence models, and lexicon-based approaches, even more flexible, consistent, and highly accurate solutions can be achieved.

Topic BERT-BiGRU is considered the most effective model for short texts. **DistilBERT**, on the other hand, excels in special cases due to its use of probability and rule-based mechanisms.

BERT + Random Forest stands out as an effective approach for integrating the model with classical machine learning methods.

CONCLUSION

This research deeply examined the process of text analysis based on the BERT model, a widely used modern approach in natural language processing. The primary advantage of the BERT model is its ability to learn context in both directions simultaneously, which allows for more precise analysis of relationships between linguistic units. This feature enables the model to work effectively with ambiguous, complex, and polysemous texts.

Throughout the study, the core architecture of the BERT model, its training phases, the role of the [CLS] token in classification, and the working mechanism of the softmax layer were analyzed. Additionally, the working mechanisms and advantages of improved models based on BERT, including DistilBERT, Topic BERT-BiGRU, and Biterm Topic Model, were thoroughly reviewed to enhance practical performance.

The lightweight and fast operation of DistilBERT was proven effective, especially in resource-limited environments. Despite having significantly fewer parameters compared to the original BERT model, it still produces results with high accuracy. The Topic BERT-BiGRU model demonstrated significantly better results by enhancing topic and semantic analysis for short texts. The Biterm Topic Model stands out for its stability when working with short and abstract texts. In particular, this model proved effective when integrated with BERT, especially for topic identification and sentiment analysis.

Recommendations:

1. Models based on the BERT model should be applied in systems such as sentiment analysis, recommendation systems, and automatic label classification to enhance efficiency.
2. If computer resources are limited, it is recommended to use DistilBERT or other lightweight BERT variants.
3. Hybrid approaches can be developed by combining the BERT model with BiGRU or CNN models for topic-based analysis or multi-class classification.
4. If there are small-sized training datasets, it is advisable to apply a **transfer learning** strategy.

Future Research Directions:

- Expanding analysis of Uzbek texts using multilingual BERT.
- Improving efficiency by fine-tuning and adapting BERT models for specific industry domains.
- Developing understandable interpretation methods for visual analysis of model results.

REFERENCES

1. Eshankulov, H.I., & Soliyeva, D.A. (2025). Natural Language Processing Models and Their Applications. *Scientific Bulletin of Bukhara State University*, 2(119), 140–146.

2. Eshankulov, K., Sayidova, N., Zaripova, G., Imomova, S., Fayzieva, D., "Mathematical Model for Information Monitoring System of Fat and Oil Enterprises," AIP Conference Proceedings, 3004(1), 060009, 2024.
3. Eshankulov, K., Murodova, R., "Development of the Knowledge Base of the Decision-Making Software Module Based on the Frame Model," Proceedings of SPIE - The International Society for Optical Engineering, 2024.
4. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is All You Need. Advances in Neural Information Processing Systems (NeurIPS). https://arxiv.org/abs/1706.03762
5. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL). https://arxiv.org/abs/1810.04805
6. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems (NeurIPS). https://arxiv.org/abs/2005.14165
7. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners. OpenAI.
8. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT). https://dl.acm.org/doi/10.1145/3442188.3445922
9. Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. (2016). Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. Advances in Neural Information Processing Systems (NeurIPS). https://arxiv.org/abs/1607.06520
10. OpenAI. (2023). GPT-4 Technical Report. OpenAI. https://openai.com/research/gpt-4