

BERT Algorithm used in Google Search

Sumeshwar Singh

Asst. Professor, Department of Comp. Sc. & Info. Tech., Graphic Era Hill University,
Dehradun, Uttarakhand India 248002

Article Info

Page Number: 1641 - 1650

Publication Issue:

Vol 70 No. 2 (2021)

Abstract

Search engines are now a need for obtaining information due to the internet's explosive expansion in digital material. One of the most widely used search engines, Google, works hard to improve its search functionality. Google has recently used cutting-edge natural language processing (NLP) methods to enhance search results. The Bidirectional Encoder Representations from Transformers (BERT) method is one such ground-breaking invention. This study seeks to offer a thorough evaluation of the BERT algorithm and its use in Google Search. We examine BERT's design, training procedure, and salient characteristics, emphasising its capacity to comprehend the subtleties and context of real language. We also talk about BERT's effects on user experience and search engine optimisation (SEO), as well as potential future advances and difficulties.

Article History

Article Received: 05 September 2021

Revised: 09 October 2021

Accepted: 22 November 2021

Publication: 26 December 2021

Keywords. BERT model, Bidirectional Encoder Representations from Transformers, Natural language understanding, Contextual understanding, Language modeling

I. Introduction

The introduction of the internet has completely changed how we access and use information. Users can now rapidly access relevant and trustworthy material thanks to search engines, which have evolved as essential tools for navigating the huge digital universe. Google has retained its supremacy among the many search engines accessible because to its ongoing efforts to enhance search quality and user experience [1]. Google has considerably improved its search skills in recent years by utilising the strength of cutting-edge natural language processing (NLP) methods, notably the Bidirectional Encoder Representations from Transformers (BERT) algorithm.

1.1 Background:

To close the gap between user intent and search results, search engines like Google face a significant difficulty. Users frequently use natural language when entering their search terms in the hopes that search engines would comprehend their purpose and provide appropriate results that meet their needs. Traditional algorithms, on the other hand, had trouble understanding the nuances of human language, which led to erroneous and inadequate search results [2].

1.2 Problem Statement:

In the parts that follow, we will present a thorough review of Google Search, its development, and the difficulties that search engine technology is now facing. The BERT algorithm [3] will next be introduced, with its neural network design, pre-training, and fine-tuning procedures explained. We

will highlight the ability of BERT to understand the subtleties of natural language through bidirectional contextual awareness.

Additionally, we will look at how BERT has affected Google Search [4], emphasising its capacity to increase contextual understanding of questions, raise the calibre of search results, enable semantic search, and broaden the capabilities of multilingual search. We will also go through how BERT may affect SEO procedures, such as how to optimise content for BERT and how it may affect keyword research and ranking aspects.

We will examine how BERT contributes to improved natural language queries, voice search, virtual assistants, and personalised search results because user experience is a key component of search engine technology. We'll also talk about BERT's drawbacks and shortcomings as well as prospective advancements in language models and search algorithms in the future.

II. Overview of Google Search:

2.1 Evolution of Search Algorithms:

Google's algorithms have advanced throughout time, moving from simple keyword matching to more complex ones that take a variety of criteria into account when delivering appropriate search results. By classifying online sites according to the quantity and calibre of links leading to them, the original PageRank algorithm revolutionised search [5]. This method gave insightful information on the authority and popularity of web pages.

Google added increasingly sophisticated algorithms as the internet expanded, such as the Panda update, which sought to raise the calibre of search results by shaming spammy and low-quality websites. With the Penguin update, manipulative link-building techniques were targeted, making search results more dependable and trustworthy. These changes demonstrated Google's dedication to improving user experience and supporting top-notch content.

2.2 Challenges in Search Engine Technology:

Despite these developments, there are still a number of difficulties that search engines must overcome in order to fully comprehend user intent and provide highly relevant search results. The ambiguity and complexity of natural language questions is one of the main difficulties [6]. Users frequently communicate their requests in conversational or informal language, making it challenging for conventional algorithms to correctly interpret their purpose.

The absence of contextual awareness presents another difficulty. Many questions include words or phrases that, depending on the context, have numerous possible interpretations. Traditional algorithms had trouble understanding the intended meaning, which led to search results that fell short of what users had come to anticipate.

Information overload is another result of the exponential increase of digital content. To guarantee that consumers obtain the most precise and fast results, search engines must effectively index and retrieve pertinent information from enormous volumes of data.

2.3 Role of Natural Language Processing:

Google has used natural language processing (NLP) methods to solve these issues. Computers can now read, interpret, and produce human language thanks to computational linguistics and machine learning, or NLP. Search engines may better understand the context, semantics, and purpose of user searches by utilising NLP [7], which produces more accurate and pertinent search results.

The BERT algorithm has had a big influence on Google Search and is a notable advancement in NLP. Search results have been revolutionised by BERT's capacity to comprehend the linguistic intricacies of context, giving consumers access to more accurate and insightful data.

III. Introducing BERT:

3.1 Understanding Neural Networks:

It is crucial to comprehend the fundamentals of neural networks before going into the BERT algorithm's details [8]. A group of machine learning models known as "neural networks" are modelled after the structure and operation of the human brain. They are made up of linked neurons arranged in layers. Up until the final output is reached, each neuron applies a mathematical operation to its inputs before passing the result to the following layer.

3.2 Transformer Architecture:

The Transformer architecture, which has become a potent framework for several natural language processing problems, is the foundation upon which the BERT algorithm is built. By relying on self-attention processes, the Transformer design enables the model to accurately capture contextual relationships and assess the relative weights of various words inside a phrase.

The concept of attention serves as the foundation for the Transformer architecture [9]. The model can concentrate on pertinent portions of the input sequence as it processes each word thanks to attention mechanisms. Due to the bidirectional nature of this attention mechanism, the model is able to better grasp context by taking into account both words that come before and after another.

3.3 Pre-training and Fine-tuning:

Pre-training and fine-tuning are the two key phases of the BERT algorithm's training process. The model learns broad language representations during pre-training by being trained on a sizable corpus of unlabeled text, such as books or articles. The model gains a richer comprehension of language structure and meaning thanks to this unsupervised learning [10].

Using a method known as masked language modelling, BERT predicts missing words in phrases during the pre-training phase. It randomly removes certain words from the input and trains the model to forecast the removed words based on the context the remaining words offer. In addition, BERT undertakes a job called next sentence prediction in which it foretells whether two phrases will come one after the other in the original text.

Following pre-training, the BERT model is refined using labelled data for a variety of downstream tasks, such as sentiment analysis or question answering. By adding more task-specific layers to the

BERT model while leaving the pre-trained weights unchanged, fine-tuning adjusts it to the particular job. Through this procedure, BERT is able to use its contextual understanding skills for a variety of NLP applications.

3.4 Bidirectional Contextual Understanding:

The BERT algorithm's aptitude for comprehending linguistic subtlety in context is one of its main advantages. By utilising its bidirectional self-attention mechanism, BERT evaluates the complete context of a phrase as opposed to conventional algorithms that just analyse words sequentially [11]. BERT can recognise cross-sentence relationships between words thanks to this bidirectional knowledge, leading to a deeper comprehension of the context.

BERT models are capable of deriving the meaning of a word from both the words that come before and after it. The constraints of earlier models, which could only access preceding words during training, are solved by BERT thanks to this bidirectional contextual awareness. As a consequence, BERT is capable of handling challenging language tasks including resolving word ambiguities and grasping sentence-level semantics that call for a deeper comprehension of context.

IV. BERT's Impact on Google Search:

4.1 Contextual Understanding of Queries:

The increased contextual comprehension of user queries provided by the BERT algorithm in Google Search is one of its main advantages [12]. Traditional search engines frequently depended on keyword matching, leaving little room for human intent to be understood. However, BERT can read the subtleties of natural language inquiries more precisely due to its bidirectional contextual awareness.

BERT can recognise the connections between words and comprehend the minute details that affect the intended meaning by taking into account the complete context of a question. BERT can give more precise and relevant search results that are in line with the user's actual intent because to this contextual awareness.

4.2 Improved Search Results:

BERT has greatly improved the accuracy and usefulness of Google search results. The algorithm can produce more accurate results since it can comprehend the context of a question. Search results produced by BERT are more likely to be in line with the user's purpose since it can handle complicated searches including numerous entities, prepositions, or other language subtleties [13].

Furthermore, BERT's bidirectional comprehension enables it to decipher lengthier searches with numerous clauses, ensuring that the search results accurately reflect the intended meaning. The shortcomings of earlier algorithms, which frequently failed to provide pertinent answers for lengthier and more complicated searches, are solved in part by this contextual knowledge.

BERT has enabled improvements in semantic search [14], which concentrates on comprehending the meaning of words and phrases in a query rather than depending just on keyword matching. Even

if the search doesn't exactly match the desired keywords, Google may still deliver search results that are in line with the user's underlying purpose because to BERT's contextual understanding.

BERT has also enhanced query expansion, allowing the computer to recognise similar ideas and broaden the search query to include more pertinent terms. This makes it possible for Google to provide search results that are more thorough, varied, and cover a wider range of relevant subjects.

4.4 Multilingual Search Capabilities:

BERT's enhanced performance in multilingual environments has a substantial influence on Google Search as well. BERT models can successfully interpret and process queries in different languages since they have been trained on vast multilingual datasets. Users who search in languages other than English now have better search experiences thanks to this development.

Language barriers are overcome by BERT's bidirectional contextual awareness, which contributes to the delivery of more precise and pertinent search results in several languages. By delivering a more inclusive and localised search experience, this innovation helps users everywhere.

V. BERT and Search Engine Optimization (SEO):

5.1 Importance of SEO in the BERT Era:

Search engine optimization (SEO) plays a crucial role in improving a website's visibility and organic traffic. With the advent of the BERT algorithm, SEO strategies have evolved to adapt to the enhanced contextual understanding of search queries. BERT's ability to comprehend the meaning behind queries and deliver more relevant search results has necessitated a shift in SEO practices.

In the BERT era, it is crucial for website owners and SEO professionals to focus on creating high-quality, informative, and user-centric content. BERT aims to provide users with the most relevant information based on their intent, rather than relying solely on keyword matching. Therefore, optimizing content for user intent, semantic relevance, and comprehensive coverage of topics has become paramount.

5.2 Optimizing Content for BERT:

To optimize content for BERT, it is important to consider the following practices:

- a) Natural Language: Write content in a natural, conversational tone that aligns with how users express their queries. BERT understands language nuances, so using natural language in your content can improve its relevance and visibility in search results.
- b) User Intent: Understand the underlying intent behind user queries and align your content to address those intents. BERT focuses on matching the user's intent rather than exact keyword matches, so tailoring your content to meet user needs is crucial.
- c) Semantic Relevance: Incorporate semantically related terms and concepts in your content to provide a more comprehensive coverage of the topic. BERT's bidirectional understanding allows it to recognize the contextual relevance of words and phrases, improving the overall semantic understanding of your content.

d) Structured Data: Utilize structured data markup, such as Schema.org, to provide additional context to search engines. Structured data helps search engines understand the content better and can enhance the visibility of your website in search results.

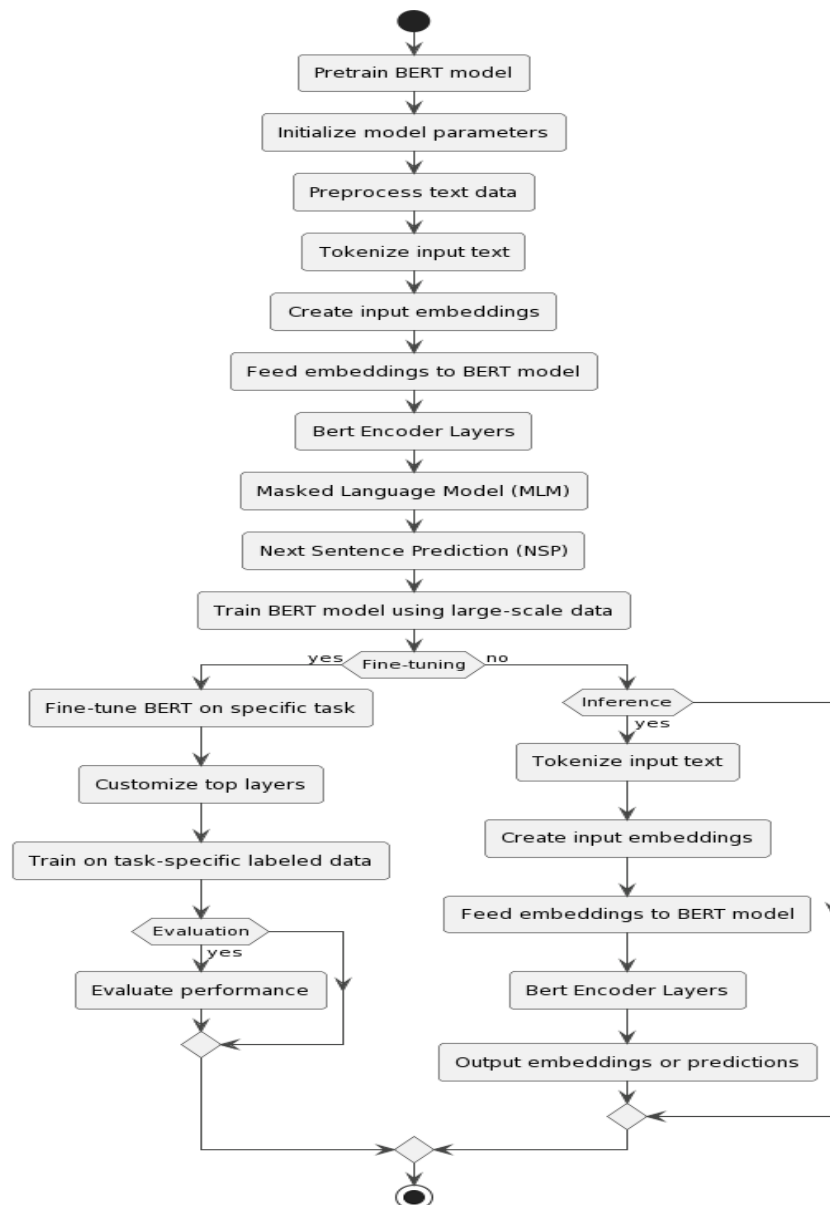


Figure.1 Working of BERT Algorithm

5.3 Impact on Keyword Research and Ranking Factors:

BERT has also influenced keyword research and ranking factors. Instead of solely focusing on exact match keywords, SEO professionals now need to consider the broader context and semantic relationships between words. Long-tail keywords and more specific phrases that reflect user intent can be more valuable in the BERT era.

Moreover, BERT's emphasis on contextual understanding has led to a shift in ranking factors. While factors like keyword relevance and backlinks remain important, search engines now prioritize factors that enhance user experience and deliver comprehensive and relevant content.

Factors such as page load speed, mobile-friendliness, user engagement, and content quality play a crucial role in achieving higher rankings.

By adapting SEO strategies to align with BERT's contextual understanding, website owners and SEO professionals can improve their website's visibility, organic traffic, and overall search performance.

VI. BERT and User Experience:

6.1 Enhanced Natural Language Queries:

BERT has greatly improved the user experience by enabling more natural language queries. Users can now express their queries in a conversational manner, using complete sentences or even fragments, and expect relevant search results. BERT's contextual understanding allows it to interpret the meaning behind the queries, considering the context and intent of the user, resulting in more accurate and helpful responses.

This enhanced natural language capability of BERT has made search interactions more intuitive and user-friendly. Users can now ask questions or describe their information needs in a more human-like manner, facilitating a smoother and more efficient search experience.

6.2 Voice Search and Virtual Assistants:

The rise of voice search and virtual assistants has been further enhanced by the BERT algorithm. Voice-based interactions rely heavily on natural language understanding, and BERT's bidirectional contextual understanding plays a key role in accurately interpreting spoken queries.

BERT helps virtual assistants, such as Google Assistant, better comprehend user commands and queries, leading to more accurate and relevant responses. Users can engage in more conversational and context-aware interactions with virtual assistants, allowing for a more seamless and personalized user experience.

6.3 Personalized Search Results:

BERT has also contributed to the personalization of search results. By understanding the context and intent behind queries, BERT enables search engines to deliver search results that are more tailored to the individual user's preferences and interests.

Personalization in search results can be achieved by taking into account factors such as a user's search history, location, and browsing behavior. BERT's ability to understand the context of queries helps in providing personalized results that align with the user's specific needs, leading to a more satisfying and relevant search experience.

6.4 Challenges and Limitations:

While BERT has made significant advancements in improving search quality and user experience, it still faces some challenges and limitations. These include:

- a) Large-scale Computation: BERT is a computationally intensive algorithm that requires substantial resources for training and inference. Implementing BERT at scale across a massive search engine like Google necessitates significant computational infrastructure.
- b) Multilingual Variations: Although BERT has shown improved performance in multilingual contexts, it still faces challenges in capturing the nuances and variations of different languages and dialects. The effectiveness of BERT may vary across languages, and ongoing research is focused on addressing these challenges.
- c) Contextual Ambiguity: While BERT excels at contextual understanding, it can still face challenges in disambiguating certain queries with complex or ambiguous meanings. Resolving these ambiguities accurately remains an active area of research.
- d) Privacy and Ethical Considerations: As search engines leverage powerful language models like BERT, ensuring user privacy and handling ethical concerns becomes crucial. Balancing the benefits of enhanced search capabilities with privacy protection and ethical considerations is an ongoing challenge.

VII. Future Developments:

The success of the BERT algorithm in enhancing Google Search has paved the way for further developments in language models and search algorithms. Some potential future directions include:

- a) Model Size and Efficiency: Research is ongoing to develop more efficient versions of BERT that can deliver similar performance with reduced computational requirements. This would enable wider adoption and deployment of BERT-like models across various platforms.
- b) Domain-Specific BERT: Adapting BERT to specific domains or industries can further enhance search relevance and precision within those domains. Customized BERT models trained on domain-specific data can better understand industry-specific terminology and context.
- c) Contextual Understanding Beyond Text: Expanding the contextual understanding capabilities of algorithms like BERT to include other modalities, such as images, audio, and video, could enable more comprehensive and multimodal search experiences.
- d) Continual Learning and Adaptation: Developing algorithms that can continuously learn and adapt to evolving user needs and preferences can further improve search experiences. This includes incorporating user feedback and interactions to refine search results over time.
- e) Multimodal Search Experiences: Integrating BERT-like models with visual and audio processing capabilities can enable more advanced multimodal search experiences. This would allow users to search using images, voice, or a combination of different modalities, expanding the possibilities of search interactions.
- f) Ethical and Fair AI: As language models like BERT become more powerful, addressing ethical considerations and ensuring fairness in search results becomes crucial. Continued research and development efforts are focused on mitigating biases, enhancing transparency, and building ethical frameworks to guide the deployment of these algorithms.

g) Advancements in NLP Techniques: Ongoing research in natural language processing is likely to lead to advancements beyond BERT. New architectures, training methods, and techniques may emerge, further enhancing the contextual understanding and performance of search algorithms.

The BERT algorithm has brought significant improvements to Google Search by enhancing contextual understanding, improving search result relevance, and providing a more intuitive and user-friendly search experience. BERT's bidirectional contextual understanding has revolutionized how search engines interpret user queries, leading to more accurate and meaningful search results. As research and development in NLP and search algorithms continue, we can expect further advancements and innovations that will shape the future of search and information retrieval.

VIII. Conclusion

Google Search has been significantly impacted by the BERT algorithm, which has completely changed how search engines interpret and handle user requests. Because of its bidirectional contextual knowledge, which permits a better grasp of language, search results are more precise and pertinent. By enabling natural language searches, enabling voice search, and providing personalised results, BERT's better contextual awareness has improved the user experience. BERT's effect affects search engine optimisation (SEO) strategies in addition to search outcomes. Instead of just depending on keyword matching, content optimisation for BERT emphasises user intent, semantic relevance, and natural language. Enhancing website visibility and organic traffic may be achieved by modifying SEO tactics to correspond with BERT's contextual understanding. Even though BERT has come a long way, there are still issues to be resolved, including the need for computing power, linguistic variances, and contextual ambiguities. The effectiveness of BERT is currently being improved, it can handle different languages, it can disambiguate complicated queries, and it can solve ethical and privacy issues. Future advancements in search algorithms and natural language processing show significant potential. More effective models, domain-specific adaptations, multimodal search experiences, ongoing learning, moral concerns, and improvements in NLP approaches are a few of these. The future of search and information retrieval will be shaped by these innovations, which will further improve the contextual comprehension and performance of search engines. The quality and relevancy of search results have generally increased greatly thanks to the BERT algorithm, providing consumers with a more user-friendly and customised search experience. BERT has established a new benchmark for search engines with its contextual awareness and capacity to grasp linguistic complexity, propelling improvements in user experience, SEO strategies, and the field of natural language processing.

References

- [1] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805. Retrieved from <https://arxiv.org/abs/1810.04805>
- [2] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2019). XLNet: Generalized autoregressive pretraining for language understanding. arXiv preprint arXiv:1906.08237. Retrieved from <https://arxiv.org/abs/1906.08237>

- [3] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. OpenAI Blog. Retrieved from <https://openai.com/blog/language-unsupervised/>
- [4] Google AI Blog. (2019). Understanding searches better than ever before. Retrieved from <https://ai.googleblog.com/2019/10/understanding-searches-better-than.html>
- [5] Schwartz, B. (2019). How BERT improves Google search results. Search Engine Land. Retrieved from <https://searchengineland.com/how-bert-improves-google-search-results-324850>
- [6] Singhal, A. (2011). Google search and search engine optimization (SEO). Google Webmaster Central Blog. Retrieved from <https://webmasters.googleblog.com/2011/05/more-guidance-on-building-high-quality.html>
- [7] Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2020). GLUE: A multi-task benchmark and analysis platform for natural language understanding. arXiv preprint arXiv:1804.07461. Retrieved from <https://arxiv.org/abs/1804.07461>
- [8] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140), 1-67. Retrieved from <http://jmlr.org/papers/v21/20-074.html>
- [9] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692. Retrieved from <https://arxiv.org/abs/1907.11692>
- [10] Google Developers. (n.d.). Search Developer Documentation. Retrieved from <https://developers.google.com/search/docs>
- [11] Li, Y., Shen, S., Ma, H., Xiao, H., & Jin, X. (2020). Deep learning for query understanding in search engines: A comprehensive survey. *ACM Transactions on Information Systems (TOIS)*, 38(4), 1-38.
- [12] Dehghani, M., Zamani, H., Severyn, A., Kamps, J., & Croft, W. B. (2017). Neural ranking models with weak supervision. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 65-74). ACM.
- [13] Yin, W., Schütze, H., Xiang, B., Zhou, B., & Zhang, B. (2019). Benchmarking neural network robustness to common corruptions and perturbations. In *Proceedings of the 36th International Conference on Machine Learning (ICML)* (Vol. 97, pp. 7204-7213). PMLR.
- [14] Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, 24(2), 8-12.
- [15] Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends in Signal Processing*, 7(3-4), 197-387.
- [16] Radinsky, K., & Horowitz, E. (2013). *The Google way: How to use Google to do everything!*. Pearson Education India.