Identifying positive and negative comments underneath technology videos based on Bert's model

Zhentao Lu[1,4], Chaoqun Qiu[2], and Zihan Yang[3]

[1]School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, China
[2]School of Integrated Circuit Science and Engineering, Nanjing University of Posts and Telecommunications, Nanjing, China
[3]School of Information Management and Artificial Intelligence, Zhejiang University of Finance and Economics, Hangzhou, China

[4]ZhentaoLu@stu.csust.edu.cn

Abstract. With the gradual popularization of the Internet and the increasing richness of content, more and more popular science videos appear in people's vision. However, people have different attitudes towards them and cannot make explicit judgments. This paper used comments from some popular science videos on Bilibili as a dataset, trained the Bert model using parallel training methods such as (Distributed Data parallel) DDP, and selected different proportions of datasets (20%, 30%, and 40%) for training and testing data to objectively determine the positive and negative attitudes of people in comments. Research has found that different scale datasets have similar performance, with an 86% -87% accuracy rate. Besides, multi-card parallel processing can bring higher returns. The results help people comprehensively understand audience feedback and emotional preferences towards technical content. At the same time, it reveals the complex relationship between the emotions reflected in comments and various factors such as audience engagement, content strategies, and demographics.

Keywords: DDP, Bert, Dataset, Training Data, Testing Data.

1. Introduction

As the Internet and social media rapidly evolve, science and technology videos have seized a formidable presence on online platforms. These videos offer a rich content menu, from debuts of novel tech products technical expositions to innovative research, appealing to a broad audience spectrum. The comment sections beneath these videos serve as vibrant communication hubs, facilitating information dissemination, opinion expression, and emotion venting. Among this diverse comment milieu are praises and approvals of the technological content, interspersed with inquiries, criticisms, and, occasionally, expressions of negative emotions. Most video platforms have deployed advanced comment moderation mechanisms to automatically identify and filter out content that contains profanity or is offensive. However, there is a category of comments that, even though they may appear linguistically correct, are intended to be negative or provocative and are often described as "weird." For tech video content publishers, advertisers, and brand promoters, it is imperative to discern these viewpoints and emotional

fluctuations. This understanding fuels insights into viewer preferences, content satisfaction, and product acceptance.

Traditional manual analysis methods grapple with the enormity of comment data. However, the advent of machine learning and natural language processing techniques, especially the pre-training techniques built on the Bert model, offers an efficacious solution [1]. The Bert model boasts considerable performance improvements across a wide range of NLP applications and is a bi-directional, pre-trained natural language processing model skilled at learning rich, context-sensitive word vector representations [2]. Chunchen Bu and other individuals created a Chinese MME dataset. To increase the precision and viability of the evaluation of troop equipment, they suggested a sentiment analysis model based on attention and contextual factors for extracting essential insights from soldiers' remarks on equipment [3]. Wenwen Zou and others proposed an external hierarchical fusion structure that fuses BERT middle layer information with non-linguistic modal multi-stage fusion for BERT fine-tuning of multimodal linguistic data on CMU-MOSI and CMU-MOSEI datasets [4]. Yusuf Ziya Poyraz explained how BERT was utilized to analyze Sindhi newspaper datasets, which addressed the difficulties in Sindhi language analysis by improving accuracy through transliteration and lexical addition [5]. Since BERT performed poorly in Chinese and needed to be improved by word-level embedding and input to improve the shortcomings of character-level coding, Lee Hao Jie and others' research sought to identify the best deep learning model for Chinese sentiment analysis, with a focus on improving the ERNIE model to improve the performance of Chinese sentiment analysis [6]. Recent research has found that pre-trained language models (e.g., BERT) perform well in NLP tasks but may suffer from learning social bias. For Chinese BERT, Sijing Zhang and others designed the first bias assessment experiment, revealing that it performs significantly on gender bias, emphasizing mitigating bias through a gender-balanced fine-tuned corpus [7].

This work uses deep learning and natural language processing techniques to analyze massive amounts of technology-focused video evaluations using the Bert model as a primary technical tool. This paper creates a manually annotated sentiment classification dataset of reviews for model training and evaluation to confirm the model's efficacy and generalizability. This methodology facilitates the identification of positive and negative tech video comments, assisting content publishers and brand promoters in deciphering viewer feedback and sentiment tendencies. The study's results will empower tech video content publishers and brand promoters to comprehend viewer feedback and sentiment fluctuations better, optimizing content strategies and communication impacts. Moreover, the findings will furnish reference points for comment section sentiment analysis applications in other domains, broadening the research's scope and depth [8-10].

2. Methods

2.1. BERT

Google's 2018 release of BERT (Bidirectional et al. from Transformers) is a ground-breaking pre-trained language model. Its novel pairing of bidirectional Training with the Transformer architecture is chiefly responsible for its game-changing impact on natural language processing [11].

Bidirectional Training. The key new feature of BERT is its capacity to consider both the left and right contexts of words inside a sentence. BERT uses bidirectional Training instead of conventional unidirectional models, which only consider words from one direction (for example, from left to right or right to left). As a result, it can capture a more thorough picture of the relationships between words in a sentence. **Transformer Architecture.** BERT is based on the Transformer architecture, a development in deep learning for natural language processing in and of itself. When formulating predictions or creating representations, the Transformer architecture relies on self-attention processes that allow it to balance the relative value of various words in a sentence. **Pre-training Phases.** The pre-training process of BERT consists of two essential phases, each contributing to its ability to learn generic language representations. **Masked Language Model (MLM).** In this stage, the Masked Language Model (MLM) is used to train BERT to predict the masked (hidden) words in a sentence. BERT learns to recognize

contextual linkages between words by hiding part of the words and making the model forecast them, which enhances its capacity to comprehend word meaning in context. The model's capacity to handle language understanding tasks depends on its performance on this MLM challenge. **Next Sentence Prediction (NSP).** Determine whether two sentences are sequential in the original text using the next sentence prediction (NSP) task. BERT is trained to recognize the logical relationship between two sentences, which aids in comprehending the coherence and flow of language. Applications involving document-level comprehension and discourse coherence can benefit most from this job.

By combining these two pre-training assignments, BERT can develop a thorough knowledge of language, including syntax, semantics, and discourse structure. The resulting model can be adjusted when used for various downstream NLP tasks, including text categorization, named entity identification, question answering, and machine translation. BERT's revolutionary approach to language modelling has paved the way for numerous advancements in natural language processing, making it a pivotal model in the field. Its ability to handle bidirectional context and effectively utilize the Transformer architecture has set new standards for NLP model development and performance.

2.2. DDP

DDP (Distributed Data Parallel) is a powerful parallelization strategy for deep learning model training in distributed computing environments. This innovative approach revolutionizes the training process by efficiently harnessing the computational power of multiple GPUs to accelerate the convergence of sizable neural network models [12].

Parallelization Strategy. DDP leverages the concept of data parallelism, which involves distributing both the model and the training data across multiple GPUs. This strategy allows concurrently processing different subsets of the training data on each GPU, effectively turning each GPU into an independent worker in the training process. **Data Distribution.** In DDP, the training dataset is divided into distinct, non-overlapping batches, each assigned to a specific GPU. This data distribution ensures that each GPU processes a unique subset of the training data. As a result, DDP can process multiple samples simultaneously, enhancing training efficiency. **Independent Computations.** Once the data is allocated to each GPU, each GPU performs forward propagation and backpropagation independently on its assigned data batch. This parallelization enables simultaneous computation of gradients for different subsets of the training data, significantly speeding up the training process for large neural networks. **Gradient Aggregation.** After computing gradients independently on each GPU, DDP employs a mechanism to aggregate these gradients. The gradients from each GPU are combined to create a global gradient update for the model parameters. This aggregation step ensures that the model parameters are synchronized across all GPUs, allowing them to contribute to model improvement collectively. **Efficient Training.** By distributing the data and computations across multiple GPUs, DDP dramatically reduces the time required for training deep neural network models. It effectively harnesses the parallel processing capabilities of GPUs, enabling faster convergence and shorter training times. **Scalability.** DDP is highly scalable, making it suitable for training models on distributed computing environments with numerous GPUs. As more GPUs are added to the training cluster, DDP can efficiently distribute the workload, maintaining training speed and efficiency.

In summary, DDP's key innovation lies in its ability to parallelize the model and the training data, enabling the concurrent processing of data batches on multiple GPUs. This parallelization accelerates the training process by performing forward and backward passes in parallel and then aggregating gradients for efficient model parameter updates. DDP has become an essential technique for training large-scale deep learning models in distributed environments, making it a crucial tool for modern machine learning research and applications. This paper focuses on training and validating a BERT-based text classifier to distinguish between positive and negative sentiments in technology video comments. The journey begins with setting up the model and deciding on the critical training parameters. These include the batch size, choice of device (GPU or CPU), the number of training iterations (epochs), and the learning rate. The careful calibration of these factors significantly dictates the course and efficiency of the training process, making them susceptible to adjustments that align with the specifics of the task.

The data collection phase involves crawling through comments on technology videos, manually tagging them, and dividing them into different proportions (20%, 30%, 40%) to serve as our training and validation data. This paper loads the pre-trained BERT model using Hugging Face's Transformers library and builds a text classification model named BertClassifier. To update the model parameters, this paper employs the AdamW optimizer, a highly effective tool in the optimization process. During the training process, this paper switches the model into training mode. This step activates specific features, such as Dropout, which aids in preventing overfitting, a common concern in model training. After processing each batch, this paper records the loss and accuracy by contrasting the predicted labels with their counterparts. After concluding each training cycle, this paper computes and documents the average loss and accuracy over the entire training dataset.

3. Experimental results and analysis

In this paper, this is achieved by choosing different proportions of the dataset, with the test set accounting for 20%, 30%, and 40% of the dataset as training and test data. Figure 1 shows the loss decrease for different dataset scales (The horizontal axis represents the number of epochs).
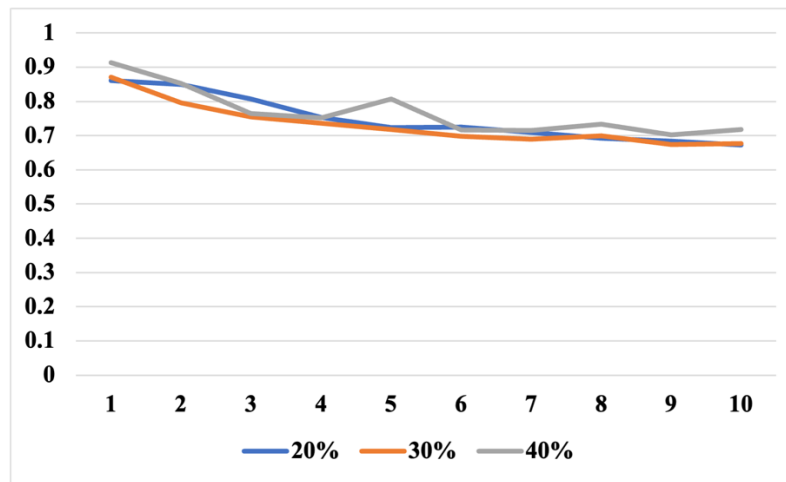
Figure 1. The loss decreases for different dataset scales.

The 20% and 30% datasets share a common characteristic as the epoch increases: the loss decreases slowly. However, on the 40% dataset, when the epoch is increased to 4, the loss decrease instead shows an upward trend, and the decrease only takes place in the fifth epoch. Similarly, at the seventh and ninth epochs, the loss decrease shows a slow upward trend, followed by a decrease. Drawing from the loss and accuracy metrics of the three experimental outcomes, the study was demonstrably successful, boasting substantial accuracy levels. The results derived from the training and testing datasets of different ratios exhibit negligible discrepancies, with the accuracy soaring as high as 86%-87% upon the execution of the 10th epoch. At the same time, the loss rate was roughly 70 percent. This paper also performs DDP processing and obtains the following experimental data, as shown in Figure 2 (The horizontal axis represents the number of graphics cards).
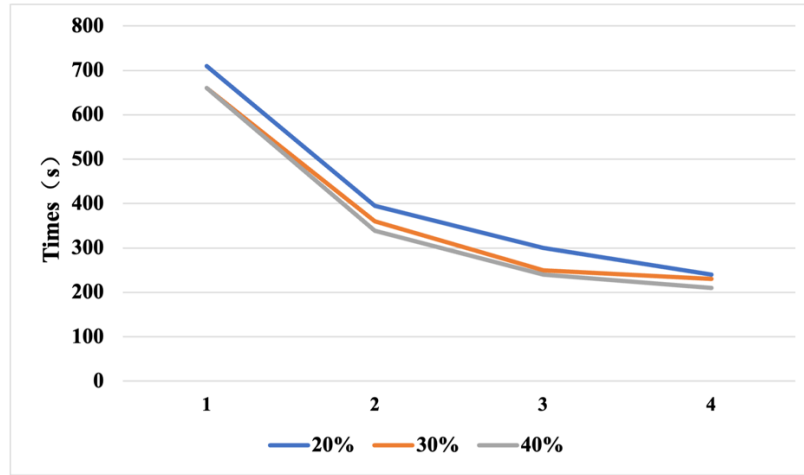
Figure 2. Experimental data under DDP parallel processing.

This paper achieves a significant speed increase with DDP, evident in the data performance. While the quad-card configuration provides the fastest speeds, the improvement from dual to triple cards is more significant than the improvement from triple to quad cards. Using three cards is the most cost-effective option. As can be seen from the data, increasing the size of the dataset only sometimes results in a linear increase in runtime. This may be because of communication between graphics cards and data distribution. This paper records the memory usage of the graphics card for each experiment under DDP parallelism to obtain Table 1.

Table 1. Graphics card memory usage per experiment under DDP parallelism.

	Dataset proportions: 20%	Dataset proportions: 30%	Dataset proportions: 40%
1 card	GPU0: 19260	GPU0: 14844	GPU0: 15834
2 cards	GPU0: 16560	GPU0: 17154	GPU0: 18594
	GPU1: 14614	GPU1: 15208	GPU1: 16648
3 cards	GPU0: 21945	GPU0: 18507	GPU0: 18507
	GPU1: 18052	GPU1: 14614	GPU1: 14614
	GPU2: 18052	GPU2: 14614	GPU2: 14614
4 cards	GPU0: 20454	GPU0: 22074	GPU0: 20454
	GPU1: 14614	GPU1: 16666	GPU1: 14614
	GPU2: 14614	GPU2: 16666	GPU2: 14614
	GPU3: 14614	GPU3: 16666	GPU3: 14614

As more GPUs are used, each GPU more evenly balances the memory load. When using Distributed Data Parallelism (DDP), video memory is only sometimes evenly distributed across GPUs. Certain overheads, such as keeping model parameters, may cause the primary GPU0 to use more memory. As the number of GPUs increases, memory allocation becomes more even. This is because of DDP's efficient memory management.

4. Conclusion
This study distinguishes between positive and negative attitudes in comments on technology videos using the BERT model as a pillar. This paper thoroughly examined the comment sections to understand the audience's opinions and emotional inclinations toward tech material. The complicated relationship between the sentiment expressed in the comments and numerous aspects, including viewer engagement, content strategy, and audience demographics, is then revealed. When used with large datasets, the BERT

model for sentiment analysis can be computationally demanding. This research used Distributed Data-Parallel (DDP) approaches to overcome the difficulties brought on by such computing requirements and guarantee an effective training process. Using DDP, a parallel training technique, the model can be trained over several GPUs, maximizing the computing capacity, and significantly lowering the training time. This paper used DDP to train the BERT model on dataset splits of 20%, 30%, and 40% for training and testing, respectively. The accuracy rate ranged from 86% to 87%, and the results were remarkably stable across these various dataset scales. This consistency highlights how vital the BERT model is when used in conjunction with the DDP technique. Additionally, the results show that using DDP to scale the training process in multi-card parallelism effectively scales the training process and ensures that the model performance is consistent and dependable across a range of dataset sizes. Thus, researchers and practitioners must use the DDP technique to train deep learning models on large datasets without sacrificing model performance or training speed. Looking ahead, this research can pave the way for further exploration into other areas, investigating the role of additional sentiment factors in social media interactions. This can provide valuable insights and practical suggestions for formulating effective content creation and dissemination strategies, creating a substantial impact across diverse content domains.

Authors Contribution
All the authors contributed equally and their names were listed in alphabetical order.

References
[1]   Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
[2]   Chen, H., Li, S., Wu, P., Yi, N., Li, S., & Huang, X. (2018). Fine-grained Sentiment Analysis of Chinese Reviews Using LSTM Network. Journal of Engineering Science & Technology Review, 11(1).
[3]   C. Bu, Y. Wang, S. Zhao, L. He and X. Wang, "Aspect-Based Sentiment Analysis of Individual Equipment Ergonomics Evaluation," 2021 International Conference on Computer, Control and Robotics (ICCCR), Shanghai, China, 2021, pp. 173-180, doi: 10.1109/ICCCR49711.2021.9349382.
[4]   W. Zou, J. Ding and C. Wang, "Utilizing BERT Intermediate Layers for Multimodal Sentiment Analysis," 2022 IEEE International Conference on Multimedia and Expo (ICME), Taipei, Taiwan, 2022, pp. 1-6, doi: 10.1109/ICME52920.2022.9860014.
[5]   Y. Z. Poyraz, M. Tugcu and M. F. Amasyali, "Improving BERT Pre-training with Hard Negative Pairs," 2022 Innovations in Intelligent Systems and Applications Conference (ASYU), Antalya, Turkey, 2022, pp. 1-6, doi: 10.1109/ASYU56188.2022.9925395.
[6]   L. H. Jie, R. K. Ayyasamy, A. Sangodiah, N. B. A. Jalil, K. Krishnan and P. Chinnasamy, "The Role of ERNIE Model in Analyzing Hotel Reviews using Chinese Sentiment Analysis," 2023 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2023, pp. 1-6, doi: 10.1109/ICCCI56745.2023.10128534.
[7]   S. Zhang and P. Li, "Unmasking the Stereotypes: Evaluating Social Biases in Chinese BERT," 2022 4th International Conference on Natural Language Processing (ICNLP), Xi'an, China, 2022, pp. 324-330, doi: 10.1109/ICNLP55136.2022.00059.
[8]   M. -Y. Day and Y. -D. Lin, "Deep Learning for Sentiment Analysis on Google Play Consumer Review," 2017 IEEE International Conference on Information Reuse and Integration (IRI), San Diego, CA, USA, 2017, pp. 382-388, doi: 10.1109/IRI.2017.79.
[9]   Tahayna, B., Ayyasamy, R. K., Akbar, R., Subri, N. F. B., & Sangodiah, A. (2022, September). Lexicon-based non-compositional multiword augmentation enriching tweet sentiment analysis. In 2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS) (pp. 19-24). IEEE.

[10] Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. IEEE access, 8, 23522-23530.

[11] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Bidirectional Encoder Representations from Transformers. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019) (Vol. 1, pp. 4171-4186).

[12] Yanping Huang, Youlong Cheng, Zenglin Xu, et al. (2019). "Distributed Data Parallel Training for Natural Language Processing." In 2019 IEEE 5th International Conference on Computer and Communications (ICCC), pp. 2223-2228.