Journal of Computer Science and Software Applications
https://www.mfacademia.org/
ISSN:2377-0430
Vol. 4, No. 3, 2024

Advancing Emotional Analysis with Large Language Models

Haowei Yang[1], Yun Zi[2], Honglin Qin[3] , Hongye Zheng[4], Yuxiang Hu[5]
[1]University of Houston ,Houston,USA
[2]Georgia Institute of Technology,Atlanta,USA
[3]Stevens Institute of Technology, Hoboken,USA
[4]Chinese University of Hong Kong, HongKong,China
[5]Johns Hopkins University,Baltimore, USA

Abstract:
The objective of this research is to enhance the efficiency of intelligence acquisition through sentiment analysis of public opinion, a crucial element of open-source intelligence, utilizing a few-shot learning framework. To address the limitations of sentiment detection models in low-resource environments due to insufficient data, we propose a novel method that integrates large language model knowledge with contrastive prompts. The methodology begins with augmenting training samples using the general knowledge from large language models. This is followed by employing unlabeled data for contrastive embedding training to improve the semantic representation capabilities of the text encoder. Finally, a prompt-based learning mechanism is used for iterative self-prediction training on the unlabeled data, further refining the model for specific tasks. Experimental results on public datasets demonstrate that the proposed model outperforms baseline methods when the same amount of labeled data is used.

Keywords:
Contrastive Learning; Prompt Learning; Large Language Model; Emotion Analysis Methods; Embedding Training

1.Introduction
Sentiment analysis[1], a quintessential task within the domain of public opinion research, is characterized by its multifaceted challenges, including the diversity of analytical needs, the complexity of textual semantics, and the timeliness required in assessments. Historically, sentiment analysis in public opinion started with dictionary-based methods, which depend on predefined lexicons of emotional words. However, these methods struggle with polysemy, negations, modifiers, and adaptability across various domains and languages. The subsequent integration of machine learning techniques marked a significant advancement, enabling models to learn classification rules from extensively annotated corpora in various fields[2-7], thereby overcoming many limitations inherent to lexicon-based approaches.
In recent years, the emergence of large-scale pre-trained language models has demonstrated substantial capabilities in language understanding[8]. The paradigm of fine-tuning pre-trained models has increasingly become a mainstream method for addressing natural language processing downstream tasks. However, this approach typically falls under supervised learning, necessitating the collection and annotation of large data samples—a process that can be resource-intensive and time-consuming, thus constraining the diversity and timeliness of public sentiment analysis.
To address these constraints, researchers have shifted their focus towards few-shot learning methodologies, which significantly reduce the need for large datasets and extensive parameter updates. Among these, prompt-based tuning is a notable few-shot learning approach that leverages the knowledge acquired during the pre-training phase, rather than merely memorizing patterns from training

data specific to a task. Despite its advantages, prompt tuning still faces challenges in ensuring accuracy when limited to single or zero-shot scenarios.

Addressing the limitations of existing studies, we propose a novel method that integrates large language models (LLMs) for initial sample annotation, coupled with unsupervised contrastive learning and unlabeled data to facilitate autonomous iterative training through prompt learning. Specifically, our approach employs LLMs to annotate samples based on sentiment classification rules, utilizes contrastive training on noised versions of the same sample to enhance the model's understanding of general patterns and features, and constructs prompt templates for iterative training on unlabeled data using the prompt learning mechanism. Experimental results indicate that our weakly-supervised method, leveraging the knowledge from large models, significantly outperforms previous models. This paper will detail our innovative methodology, the integration of LLMs with contrastive prompts, and how these techniques collectively refine the accuracy and applicability of sentiment analysis for intelligence acquisition in a variety of settings.

2.Related work

The landscape of emotional analysis within public opinion research has experienced a remarkable evolution, significantly influenced by advances in machine learning and language modeling. Deep learning has achieved significant advancements in image segmentation[9-11] and medical diagnosis[12-15] in recent years, with notable performance in sentiment analysis. Neural network models like CNN [16], BiGRU [17], and LSTM [18-19] have been incorporated into sentiment analysis. This progression is illustrated by recent literature that focuses on the effective deployment of large-scale pre-trained language models (LLMs) and innovative learning methodologies. For instance, Mei et al. [20]have optimized LLMs using deep learning techniques to enhance performance on complex NLP tasks, a strategy that supports our approach of integrating LLMs for initial sentiment analysis and highlights the synergy between large model capabilities and specialized tasks such as sentiment classification. Similarly, Xu et al. [21] addressed the challenges of data-intensive training by optimizing LSTM networks for financial risk prediction, which parallels our use of prompt-based tuning for efficient learning, aimed at overcoming data scarcity constraints. In the realm of medical informatics, Xiao et al.[22] and Wang et al[23].explored attention mechanisms to better extract insights from medical texts, underscoring the potential of advanced neural networks in handling complex data, which is relevant to our application of these principles for nuanced textual semantics in public sentiment. Additionally, Chai et al. [24]and Gao et al. [25] extended these methodologies to image data, applying deep learning for lung image recognition and enhancing encoder-decoder architectures for semantic segmentation, respectively. These studies demonstrate the utility of deep learning in analyzing complex patterns and underscore our use of contrastive learning strategies to refine the model's accuracy and robustness by training on noised text samples, thereby enhancing sentiment classification in public opinion research.

3.Methodology

3.1 Few-shot learning

Few-shot learning is a machine learning framework designed to teach models to generalize effectively from a very limited number of labeled examples, typically in the order of one to five samples per class. This approach is crucial in scenarios where data acquisition and annotation are expensive or impractical[26].

In the academic context, few-shot learning is often formalized using the concept of episodic training, where the model is exposed to a series of N-way, K-shot learning tasks[27]. In this setting, N represents the number of classes, and K signifies the number of examples per class available for training. The objective for the model in each episode is to learn to classify correctly using only these K examples[28]. Mathematically, the training process in a few-shot learning scenario can be represented by:

$$\text{minimize} \quad E_{(S,Q)\sim\mathbb{D}}[\mathcal{L}(f_\theta(S), Q)] \tag{1}$$

where S (the support set) consists of K samples for each of the N classes, Q (the query set) contains new, unseen examples from the same N classes for testing, D is the distribution over tasks, f_θ is the predictive function parameterized by θ, and L is a loss function measuring the prediction error on the query set.

Furthermore, meta-learning or "learning to learn" approaches are often employed within few-shot learning. These methods train a meta-learner on a variety of learning tasks so that it can quickly adapt to new tasks with only a few training examples. One popular meta-learning technique is the Model-Agnostic Meta-Learning (MAML) algorithm, which optimizes a model's initial parameters so that a few gradient updates will lead to good generalization on new tasks. The MAML update rule is expressed as:

$$\theta' = \theta - \alpha \nabla_\theta \mathcal{L}train(f_\theta) \tag{2}$$

where θ, represents the adapted parameters after training on the support set of a new task, α is the learning rate, and $\nabla_\theta \mathcal{L}train$ is the gradient of the training loss with respect to the parameters θ.

These techniques underscore the effectiveness of few-shot learning in tackling the challenges of data scarcity, allowing models to apply learned knowledge to new tasks with minimal additional data.

3.2 Model Architecture

This paper introduces a novel sentiment analysis method based on contrastive sentiment analysis with large language model knowledge (CSLLM). As illustrated in Figure 1, the primary components of this model are unsupervised contrastive embedding training, data annotation using a large language model, initial model training, and unsupervised iterative training. The proposed approach leverages the strengths of large language models in understanding and generating human-like text, thereby enhancing the accuracy and robustness of sentiment analysis tasks. By incorporating unsupervised learning techniques, the model can effectively handle vast amounts of unlabelled data, making it highly scalable and adaptable to various domains.

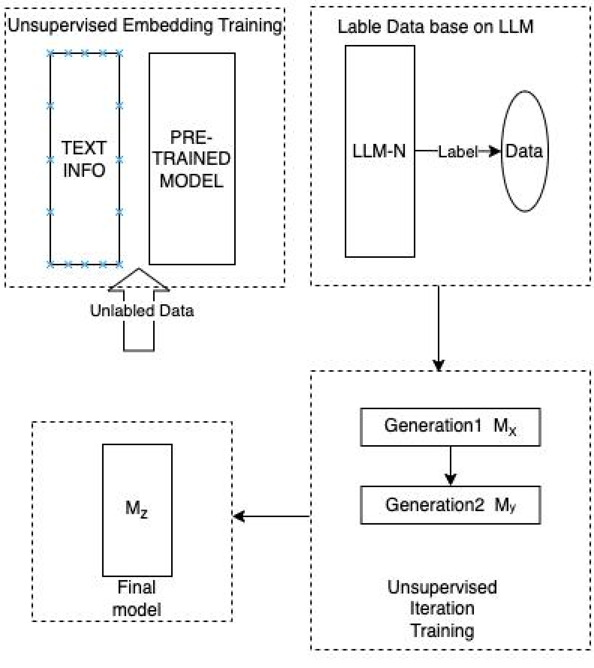

Figure 1. CSLLM Architecture

3.3 Text Annotation Based on Large Language Models

To implement text annotation based on large language models, especially when training the initial generation of the model, the training data might be minimal or completely unavailable. In these circumstances, constructing an initial annotated dataset is achieved by designing suitable question-and-answer templates and leveraging large language models. By employing multiple large language models for text annotation, we select samples with consistent annotations as the initial training dataset, limiting to a maximum of 100 samples annotated by the models. Considering the varying levels of human annotation in the dataset, we devised the following three annotation strategies: 1-shot: Provide one manually labeled training sample for each category. With necessary instructions, employ the large language model to generate 100 annotated samples, and select samples with consistent labels to form the initial dataset. 50-shot: Building on the 1-shot approach, an additional 50 manually labeled samples are included. 100-shot: Manually label 100 samples to create the initial training set without utilizing any large language models.

4.Experimental Design and Results

To validate the effectiveness of the proposed few-shot public opinion analysis method, we conducted an empirical study. This section compares the sentiment analysis capabilities of the CSLLM model with those of a baseline model.

4.1 Dataset and Prompt Templates

We selected two typical sentiment analysis tasks from the publicly available TweetEval [29] Twitter dataset for experimental analysis: discrimination speech detection and attack speech detection. For these tasks, we designed corresponding prompt templates. By filling the samples into the templates, sentiment analysis is achieved using a text completion approach. The dataset uses Linked Data methodology to merge diverse data formats, improving interoperability and analysis in machine learning and AI. This method breaks down data silos, enriches dataset diversity, and supports more accurate and scalable AI applications.[30]

The prompt templates were constructed using both direct prompts and context-enriched prompts: with direct prompts, the model needs to directly understand the text and task to perform sentiment analysis. Context-enriched prompts, on the other hand, first provide a brief task introduction before engaging in understanding and classification.

4.2 Evaluation Metrics

The primary metric used to evaluate the model performance is accuracy. Additionally, the F1-score is incorporated to assess the large language model’s capacity to handle data with imbalanced positive and negative examples. For sentiment analysis, the possible combinations of model predictions and actual sentiments are defined as follows: True Positive (TP): the model predicts a positive sentiment, and it is indeed positive; False Negative (FN): the model predicts a negative sentiment, but it is actually positive; True Negative (TN): the model predicts a negative sentiment, and it is indeed negative; False Positive (FP): the model predicts a positive sentiment, but it is actually negative. Based on these definitions, we have:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

$$P = \frac{TP}{TP+FP},\ R = \frac{TP}{TP+FN},\ F1 = 2 \times \frac{P \times R}{P+R} \tag{4}$$

4.3 Baseline Models

To compare with CSLLM under identical resource conditions, the following baseline methods were selected:

1) BERT [31]: A transformer-based bidirectional encoder representation model that exhibits robust performance in various natural language processing tasks.

2) RoBERTa [32]: An enhanced version of BERT with improved pre-training strategies and additional parameters. The RoBERTa-Base variant was used for comparison.
3) XLNet [33]: A transformer model utilizing an autoregressive pre-training approach, outperforming BERT in several benchmarks.
4) PET-{BERT, GPT-2}: A prompt-based learning method that employs Pattern-Exploiting Training (PET) using BERT/GPT-2 models.
5) iPET-{BERT, GPT-2}: This method uses iterative Pattern-Exploiting Training (iPET) based on BERT/GPT-2 models. iPET maximizes the utilization of pre-trained model knowledge through iterative training on unlabeled samples.

The training resources available for these baseline methods are as follows:

6) 1-shot: One manually labeled training sample per category.
7) 50-shot: A total of 50 manually labeled training samples.
8) 100-shot: A total of 100 manually labeled training samples.

For sample annotation, the following large language models were utilized:

GPT-3.5-Turbo: An upgraded version of GPT-3 with enhanced features and parameters, achieving superior performance in open-domain QA tasks.ChatGLM-6B: A bilingual model based on the General Language Model framework with 6.2 billion parameters. This model can be deployed on personal devices using quantization techniques.

Additionally, the text-Davinci-002 model was introduced as a control for sample annotation, showcasing excellent performance in text reasoning due to its fine-tuning with instructions.

4.4 Experimental Setup

For the CPLLA-SA model, the temperature hyperparameter and dropout rate for contrastive embedding training were set to 0.06 and 0.1, respectively. The auxiliary loss weight was set to 1e-4. The model's maximum participation rate was set to 0.52, the generational sample increment multiplier d was set to 5, and the iterative selected sample size w was set to 10. For GPT-3.5-Turbo, the hyperparameter was set to 0.3. The Chat-GLM model used was a bit-4 quantized version with default parameter settings. For the PET model, "reduction" was set to "mean" and "ipet_n_most_likely" was set to 10.

Given that TweetEval is a short text classification dataset, the maximum text read length for the experimental models was set to 128, the training batch size was set to 7, and the training epochs ranged from 4 to 6, depending on model performance. The learning rate for prompt fine-tuning was 5e-4, while for conventional fine-tuning it was 2e-4.

The experiments were conducted on a platform running Windows 11, utilizing the deep learning libraries torch and transformers 4.23.0 with Python 3.9.12. The hardware configuration included an AMD Ryzen 7 5800H CPU, NVIDIA GeForce RTX 3070 Laptop GPU, a 2TB SSD, and 32GB of RAM.

4.5 Analysis of Experimental Results

Based on the TweetEval dataset, experiments were conducted for discrimination speech detection and attack speech detection. Accuracy was used as the evaluation metric. The results are presented in Tables 1 and 2.

Table 1 Discrimination Speech Detection Results

Model	1-shot	50-shot	100-shot
BERT	0.4969±0.0323	0.6082±0.0297	0.6402±0.0265
RoBERTa	0.4756±0.0687	0.5729±0.0000	0.6502±0.0157
XLNet	0.4622±0.0432	0.5872±0.0039	0.5932±0.0101
PET-GPT2	0.4527±0.0163	0.5942±0.0074	0.6114±0.0121
PET-BERT	0.4334±0.0116	0.5889±0.0468	0.6449±0.0091

Journal of Computer Science and Software Applications
https://www.mfacademia.org/
ISSN:2377-0430
Vol. 4, No. 3, 2024

iPET-GPT2	0.5247±0.0546	0.6049±0.0070	0.6422±0.0143
iPET-BERT	0.5192±0.0897	0.6354±0.0032	0.6604±0.0036
CPLLM-SA	0.6332±0.0111	0.6549±0.0137	0.6662±0.0075

Table 2 Attack Speech Detection Results

Model	1-shot	50-shot	100-shot
BERT	0.4878±0.0133	0.5196±0.0367	0.5727±0.0159
RoBERTa	0.4928±0.0187	0.4774±0.0000	0.5447±0.0170
XLNet	0.4791±0.0101	0.5224±0.0064	0.5545±0.0132
PET-GPT2	0.5001±0.0386	0.5201±0.0479	0.5795±0.0464
PET-BERT	0.4811±0.0074	0.5075±0.0183	0.5410±0.0122
iPET-GPT2	0.5258±0.0490	0.5290±0.0182	0.5784±0.0133
iPET-BERT	0.5030±0.0359	0.5321±0.0135	0.5622±0.0032
CPLLM-SA	0.5070±0.0277	0.5687±0.0112	0.5842±0.0156

From the experimental results, it can be observed that in the discrimination speech detection task, the accuracy of the CSLLM model improved by approximately 11 percentage points over the best baseline method, iPETBERT, in the 1-shot training scenario. In the 50-shot and 100-shot training scenarios, the accuracy increased by about 2.0 and 0.6 percentage points, respectively. In the attack speech detection task, the CSLLM model underperformed compared to the iPET-GPT2 method in the 1-shot scenario, suggesting that large language models struggle with implicitly expressed short texts. However, with an increase in the number of real samples, the CSLLM model achieved the best performance in the 50-shot and 100-shot scenarios, with accuracy improvements of approximately 3.7 and 2.1 percentage points over the second-best model.

5. Conclusion

The research highlighted in this paper offers a substantial enhancement to sentiment analysis techniques, particularly in low-resource environments. By integrating the advanced capabilities of large language models (LLMs) with novel methodologies like prompt learning and contrastive sentiment analysis with large language model knowledge (CSLLM). not only addresses the challenge of data scarcity but also sets a new standard for performance in sentiment analysis.Our approach effectively minimizes the need for extensive labeled datasets, which are often a barrier in traditional sentiment analysis. The experimental results demonstrate that CPLMM-SA excels in detecting complex emotional nuances in text, even when limited labeled data is available. This advantage is particularly significant for applications in open-source intelligence, where rapid and accurate sentiment assessment is crucial.Future work will explore the integration of online knowledge graphs to enable more dynamic, zero-shot sentiment analysis. This development promises to enhance the model's responsiveness to evolving scenarios and its ability to understand sentiment without prior specific training. Additionally, further refining the contrastive and prompt-based techniques will allow our methodology to be effectively applied across different languages and cultural contexts, broadening its utility and impact.In conclusion, this approach represents a promising development in the field of sentiment analysis, pushing the boundaries of what is achievable with current technology and providing a robust foundation for future advancements.

References

[1] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. Ain Shams engineering journal, 5(4), 1093-1113.

[2] Zhao, B., Cao, Z., & Wang, S. (2017). Lung vessel segmentation based on random forests. Electronics Letters, 53(4), 220-222.

[3] Liu, Z., & Song, J. (2021, November). Comparison of Tree-based Feature Selection Algorithms on Biological Omics Dataset. In Proceedings of the 5th International Conference on Advances in Artificial Intelligence (pp. 165-169).

[4] Li, M., Zhu, Z., Xu, R., Feng, Y., & Xiao, L. (2024). Research on Image Classification And Semantic Segmentation Model Based on Convolutional Neural Network. Journal of Computing and Electronic Information Management, 12(3), 94-100.

[5] Lin, T., & Cao, J. (2020). Touch interactive system design with intelligent vase of psychotherapy for alzheimer’ s disease. Designs, 4(3), 28.

[6] Zhang, H., Diao, S., Yang, Y., Zhong, J., & Yan, Y. (2024). Multi-scale image recognition strategy based on convolutional neural network. Journal of Computing and Electronic Information Management, 12(3), 107-113.

[7] Yang, Y., Chen, Z., Yan, Y., Li, M., & Gegen, T. (2024). A new method of image recognition based on deep learning generates adversarial networks and integrates traditional algorithms. Journal of Computing and Electronic Information Management, 13(1), 57-61.

[8] Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., ... & Sun, M. (2023). Parameter-efficient fine-tuning of large-scale pre-trained language models. Nature Machine Intelligence, 5(3), 220-235.

[9] Xu, T., Li, I., Zhan, Q., Hu, Y., & Yang, H. (2024). Research on Intelligent System of Multimodal Deep Learning in Image Recognition. Journal of Computing and Electronic Information Management, 12(3), 79-83.

[10] Yuan, J., Wu, L., Gong, Y., Yu, Z., Liu, Z., & He, S. (2024). Research on Intelligent Aided Diagnosis System of Medical Image Based on Computer Deep Learning. arXiv preprint arXiv:2404.18419.

[11] Wang, Q., Schindler, S. E., Chen, G., Mckay, N. S., McCullough, A., Flores, S., ... & Benzinger, T. L. (2024). Investigating White Matter Neuroinflammation in Alzheimer Disease Using Diffusion-Based Neuroinflammation Imaging. Neurology, 102(4), e208013.

[12] Yan, Y., He, S., Yu, Z., Yuan, J., Liu, Z., & Chen, Y. (2024). Investigation of Customized Medical Decision Algorithms Utilizing Graph Neural Networks. arXiv preprint arXiv:2405.17460.

[13] Lu, S., Liu, Z., Liu, T., & Zhou, W. (2023). Scaling-up medical vision-and-language representation learning with federated learning. Engineering Applications of Artificial Intelligence, 126, 107037.

[14] Yan, X., Wang, W., Xiao, M., Li, Y., & Gao, M. (2024). Survival Prediction Across Diverse Cancer Types Using Neural Networks. arXiv preprint arXiv:2404.08713.

[15] Yao, Z., Lin, F., Chai, S., He, W., Dai, L., & Fei, X. (2024). Integrating Medical Imaging and Clinical Reports Using Multimodal Deep Learning for Advanced Disease Analysis. arXiv preprint arXiv:2405.17459.

[16] Chua, L. O., & Roska, T. (1993). The CNN paradigm. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 40(3), 147-156.

[17] Bano, S., Khalid, S., Tairan, N. M., Shah, H., & Khattak, H. A. (2023). Summarization of scholarly articles using BERT and BiGRU: Deep learning-based extractive approach. Journal of King Saud University-Computer and Information Sciences, 35(9), 101739.

[18] Dai, W., Tao, J., Yan, X., Feng, Z., & Chen, J. (2023, November). Addressing Unintended Bias in Toxicity Detection: An LSTM and Attention-Based Approach. In 2023 5th International Conference

on Artificial Intelligence and Computer Applications (ICAICA) (pp. 375-379). IEEE.

[19] Zhan, Q., Ma, Y., Gao, E., Sun, D., & Yang, H. (2024). Innovations in Time Related Expression Recognition Using LSTM Networks. International Journal of Innovative Research in Computer Science & Technology, 12(3), 120-125.

[20] Mei, T., Zi, Y., Cheng, X., Gao, Z., Wang, Q., & Yang, H. (2024). Efficiency optimization of large-scale language models based on deep learning in natural language processing tasks. arXiv preprint arXiv:2405.11704.

[21] Xu, K., Cheng, Y., Long, S., Guo, J., Xiao, J., & Sun, M. (2024). Advancing Financial Risk Prediction Through Optimized LSTM Model Performance and Comparative Analysis. arXiv preprint arXiv:2405.20603.

[22] Xiao, L., Li, M., Feng, Y., Wang, M., Zhu, Z., & Chen, Z. (2024). Exploration of Attention Mechanism-Enhanced Deep Learning Models in the Mining of Medical Textual Data. arXiv preprint arXiv:2406.00016.

[23] Wang, S., Liu, Z., & Peng, B. (2023, December). A Self-training Framework for Automated Medical Report Generation. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 16443-16449).

[24] Chai, S., Fei, X., Wang, Y., Dai, L., & Sui, M. (2024). Deep Learning-Based Lung Medical Image Recognition. International Journal of Innovative Research in Computer Science & Technology, 12(3), 100-105.

[25] Gao, Z., Wang, Q., Mei, T., Cheng, X., Zi, Y., & Yang, H. (2024). An Enhanced Encoder-Decoder Network Architecture for Reducing Information Loss in Image Semantic Segmentation. arXiv preprint arXiv:2406.01605.

[26] Liu, S., Yan, K., Qin, F., Wang, C., Ge, R., Zhang, K., ... & Cao, J. (2024). Infrared Image Super-Resolution via Lightweight Information Split Network. arXiv preprint arXiv:2405.10561.

[27] Xu, R., Yang, Y., Qiu, H., Liu, X., & Zhang, J. (2024). Research on Multimodal Generative Adversarial Networks in the Framework of Deep Learning. Journal of Computing and Electronic Information Management, 12(3), 84-88.

[28] Jiang, H., Qin, F., Cao, J., Peng, Y., & Shao, Y. (2021). Recurrent neural network from adder’s perspective: Carry-lookahead RNN. Neural Networks, 144, 297-306.

[29] Barbieri, F., Camacho-Collados, J., Neves, L., & Espinosa-Anke, L. (2020). TweetEval: Unified benchmark and comparative evaluation for tweet classification. arXiv preprint arXiv:2010.12421.

[30] Li, Y., Yan, X., Xiao, M., Wang, W., & Zhang, F. (2024). Investigation of Creating Accessibility Linked Data Based on Publicly Available Accessibility Datasets. In Proceedings of the 2023 13th International Conference on Communication and Network Security (pp. 77–81). Association for Computing Machinery.

[31] 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.

[32] 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.

[33] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. Advances in neural information processing systems, 32.