Eng Int Syst (2026) 1: 17–26

Engineering
Intelligent Systems

Exploration of Enhancement Effect in Natural Language Understanding Task Based on BERT Model with Integrated Power Knowledge Graph

Weidong Guo[1,a], Yubin Xu[1,b], Zhifeng Shi[1,c*], Xuefeng Hu[1,d], Huiping Zhang[1,e], Lei Feng[2,f]

[1] *State Grid Shanxi Electric Power Company, Taiyuan 030001, Shanxi, China*
[2] *State Grid Shanxi Electric Power Company Material Branch, Taiyuan 030001, Shanxi, China*

In response to the problems of poor domain adaptability and the weak knowledge fusion ability of traditional language models in NLU (Natural Language Understanding) tasks, which prevent them from fully capturing deep relationships in context, this studyaims to integrate power knowledge graphs to enhance the effectiveness of BERT (Bidirectional Encoder Representations from Transformers) models in NLU tasks, enabling them to more accurately infer the terminology and contextual meanings pertaining tothe electricityfield, thereby improving training efficiency and model performance, and promoting the development of automation in this field. The performance evaluation results of the BERT model integrating power knowledge graph in NLU tasks were: an average path length of 3.8, language similarity of 0.9, and vocabulary coverage of 0.8, all of which were superior to other models used for comparison. The experimental results showed that the BERT model integrating a power knowledge graph had better performance compared to other models commonly used for processing NLU tasks.

Keywords: power knowledge graph; BERT model; natural language understanding; power sector; knowledge fusion

1. INTRODUCTION

Natural Language Understanding (NLU) is a significant research area in the field of artificial intelligence (AI) that aims to enable computers to understand and interpret human language. As deep learning techniques develop, neural network-based language models, such as the BERT model, have made significant progress in various NLU tasks [1–3]. However, these general-purpose language models tend to perform poorly when dealing with domain-specific texts, mainly because of their lack of understanding and integration of domain-specific knowledge. Particularly in the highly specialized field of electricity, language models need to have a rich background of professional knowledge in order to accurately understand and reason about relevant terms and contexts.

Although the BERT model has demonstrated strong performance in natural language processing tasks, its application in the field of power still faces many challenges. Firstly, the

*Corresponding author: Address for correspondence:Zhifeng Shi, State Grid Shanxi Electric Power Company, Taiyuan 030001, Shanxi, China, E-mail: [a]guoweidong528@163.com, [b]xybing456781@yeah.net, [c]szf55551018@126.com, [d]huxuefeng0087@sina.com, [e]sjzhplz@126.com, [f] FL3211@163.com

training data for traditional BERT models are mostly generic corpora, lacking specialized terminology and knowledge for the power field, resulting in insufficient comprehension when processing text in this field. Secondly, the BERT model fails to effectively integrate the power knowledge graph (KG) during the pre-training stage, and is unable to fully utilize existing domain knowledge to enhance its understanding and reasoning ability of the text. This makes it difficult for the model to capture deep relationships and semantics when faced with complex power domain contexts. In addition, how to effectively evaluate the BERT model's performance in the NLU task after integrating the power knowledge graph is also an urgent issue.

This study addresses the aforementioned issues by constructing and integrating a power knowledge graph to improvethe effectiveness of BERT models in NLU tasks. Firstly, the method used to construct a power knowledge graph is explainedin detail, and includesdata collection and processing. Next, how to effectively integrate the power knowledge graph with the BERT model is discussed. By applying knowledge embedding techniques, the BERT model can fully utilize professional knowledge in the power field during the pre-training and fine-tuning stages. Then, a fused model framework is designed and appropriate model metrics are selected to evaluate its performance. Finally, a series of simulation experiments are conducted to verify the effectiveness of the BERT model incorporating the power knowledge graph in the NLU task, and a comparative analysis is conducted with the traditional BERT model to evaluate its enhancement effect in terms of text comprehension and reasoning ability in the electric power domain.

2. RELATED WORK

There has been a considerable amount of research grounded in the BERT model, which is utilized mainly for text classification tasks in different languages [4–6]. In order to extract important geographic application information from temporalinformation in social media messages, Ma et al. [7] used the BERT model and proposed an automatic extraction algorithm for social media message time information based on deep learning, effectively derivinggeographic application information from time information in social media messages. In response to the lack of sentiment analysis models for Hebrew language used in NLU, Chriqui et al. [8] proposed HeBERT and HebEMO for Hebrew text analysis, polarity analysis, and sentiment recognition, and demonstrated their feasibility through research. Against the backdrop of the development of an increasing number of digital text corpora and NLP (Natural Language Processing) tools and methods, scholars such as Uveges [9] successfully created a Transformer-based fine-tuning model and used this finely-tuned BERT model to classify emotions and feelings in Hungarian political communication.

In addition, there is also a very close connection between the BERT model and NLU tasks. The BERT model can capture contextual semantics and is therefore widely used forNLU tasks, including sentiment analysis [10–12] and multi-label text classification [13–15]. In order to better enable computers to understand human language, researchers have integrated BERT models with other advanced technologies to construct new methods for different NLU tasks. Due to the tendency of BERT models to overlook the dependency relationships between predicted tokens during pre-training, and the issue of positional differences between pre-training and fine-tuning when applying XLNet, Song K et al. [16] proposed a successful new pre-training method for NLU tasks by integrating the advantages of BERT and XLNet. Wang B et al. [17] presented a new sentence-embedding method by studying the hierarchical pattern of word representations in deep contextualization models, and demonstrated its effectiveness through experiments. In response to BERT's heavy reliance on global self-attention modules, which leads to high memory usage and computational costs, Jiang Z H et al. [18] proposed a new span-based dynamic convolution to replace these self-attention heads, successfully reducing training costs and model parameters.

However, in the field of electricity, there is little research on BERT models that integrate power knowledge graphs. Although researchers are committed to building BERT models that integrate knowledge graphs [19–21], such models are still rare [22]. Therefore, this present studyconstructs a BERT model that integrates power knowledge graphs, enabling it to perform better in NLU tasks, thereby enriching the relevant research on the NLU tasks of BERT models that integrate power knowledge graphs in the electricityfield.

3. BERT MODEL INTEGRATING POWER KNOWLEDGE GRAPH

3.1 Construction of Power Knowledge Graph

A power knowledge graph is a tool that represents professional knowledge in the field of electric power by means of a graph structure, the nodes and edges of which represent entities and their relationships. The construction of a knowledge graph involves three main stages: knowledge acquisition, modeling, and graphing. These processes organize structured and unstructured information in the field of electricity into computable and searchable knowledge graphs, and provide professional knowledge in areas such as power equipment, process flow, and technical standards.

The first step in building a power knowledge graph is knowledge acquisition, which involvesextracting professional knowledge from various data sources in the power industry. To extract useful information from these unstructured data, NLP techniques such as Named Entity Recognition (NER) and relationship extraction [23–24] are used, which can automatically identify and extract important entities in text (such as transformers, generators) and their relationships (such as "transformer-transformer ratio-generator"). In the knowledge modeling stage, RDF (Resource Description Framework) and OWL (Web Ontology Language) are used to describe these entities and their relationships. Through the application of graph databases (such as Neo4j) [25], modeled knowledge is input into the graph database to form a complete

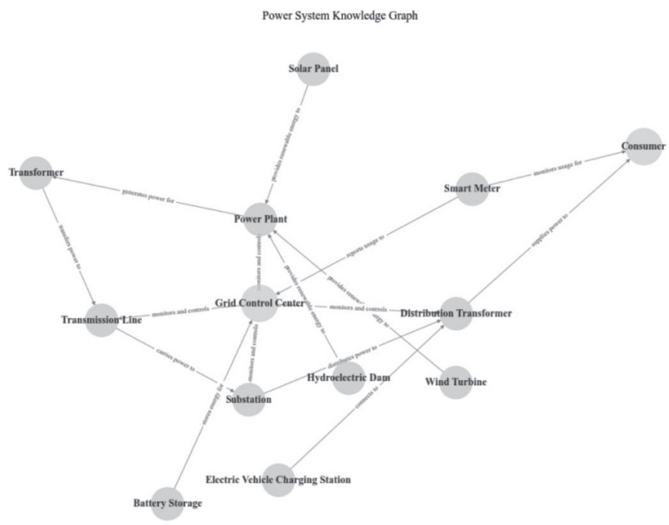

Figure 1 Diagram of a power knowledge graph.

knowledge graph, achieving efficient storage, querying, and reasoning of data.

Figure 1 depictsa power knowledge graph.

Figure 1 illustrates the interrelationships and information flow between various components ofthe power system. The main elements include power generation equipment (such as power plants, wind turbines, solar panels), transmission equipment (such as transformers, transmission lines), distribution systems (such as substations, distribution transformers), smart devices (such as smart meters, electric vehicle charging stations), and control centers. In Figure 1, arrows represent the flow relationship of electricity or information, such as power generation equipment providing energy to transmission equipment, and the control center monitoring and regulating the operation of the entire system.

3.2 Integration Mechanism of Knowledge Graph and BERT

The power knowledge graph is fused with the BERT model to enhance the adaptability and reasoning ability of the BERT model in the power domain[26]. The rich domain knowledge in the knowledge graph is utilized to compensate for BERT's shortcomings in specific professional knowledge in general corpus pre-training. Bycombining the entity and relationship information in the knowledge graph with the word vectors of the BERT model, more semantic and background knowledge can be embedded in the model, thereby enhancing its understanding of the power field [27].

The fusion mechanism of the knowledge graph and the BERT model consists of three main steps: knowledge embedding, semantic matching, and model integration. In the knowledge embedding stage, graph embedding techniques (such as TransE, TransH, and RotatE) are used to transform entities and relationships in the power knowledge graph into low dimensional vectors [28–29]. These methods embed entities and relationships into the same vector space through different mapping strategies, enabling them to be processed and optimized in neural networks. In the semantic matching stage, natural language processing techniques [30–31] are used to identify entities in the power field in the input text and align them with entities in the knowledge graph, and includeusing similarity calculation methods (such as cosine similarity) to compare the input text with entities in the graph to ensure that the model containsthe most relevant knowledge. In the model integration stage, a fusion strategy is designed to combine the embedding vectors in the knowledge graph with the word vectors of the BERT model. A knowledge enhancement layer is added to the Transformer encoder of BERT, which integrates the information in the knowledge graph with the contextual features of BERT through an attention mechanism or a gating mechanism [32–34]. This integration approach enables the BERT model to better understand the semantics and context of the power field, improving its ability to understand professional terminology and complex relationships.

3.3 Integrated Model Framework

The BERT model framework that integrates power knowledge graph has four main components: input module, encoder module, knowledge enhancement layer, and output layer. Figure 2 shows the fused model framework.

As illustrated in Figure 2, the input module is responsible for simultaneously receiving and processing text data and knowledge graph information. After the text is segmented and tokenized, the word vectors required for BERT are generated. At the same time, entities and relationships related to the input text are extracted from the power knowledge graph and converted into knowledge embedding vectors. The encoder module adopts a multi-layer Transformer architecture to encode text word vectors and knowledge graph vectors. By using a self-attention mechanism to calculate the attention weights between different modal information, important semantic relationships can be captured. Specifically, in the encoding layer of Transformer, a knowledge enhancement layer is applied to integrate the knowledge graph embedding vector with text features. The knowledge enhancement

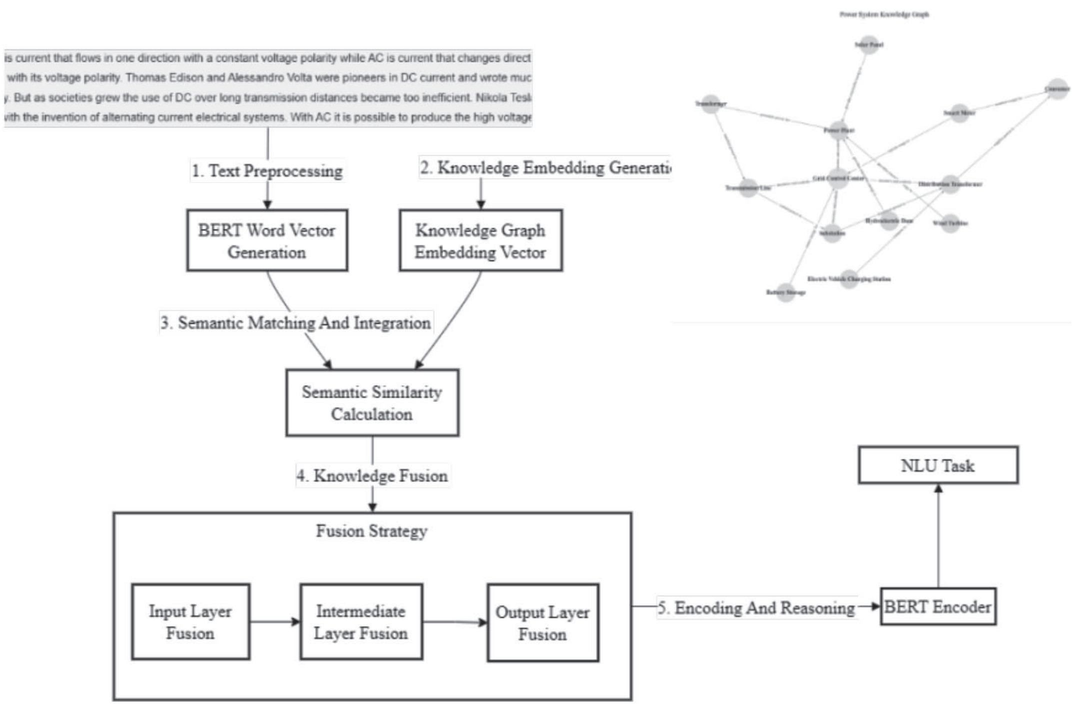

Figure 2 The fused model framework.

layer is designed to embed domain knowledge more deeply into BERT's encoder. By utilizing an attention mechanism, this layer can selectively combine information from the knowledge graph and integrate it with the text context by means ofconcatenation or weighted fusion strategies. The output layer generates the final model output based on specific task requirements, such as text classification or sequence annotation.

3.4 Model Indicator Selection

Four aspects of metrics are selected to assess the BERT model's performance incorporating the power knowledge graph in the NLU task, namely path reasoning ability, semantic understanding, professional terminology processing ability, and model accuracy and generation performance.

The indicators of path reasoning ability include the average path length λ; the indicators for semantic understanding include semantic similarity; the indicators of professional terminology processing ability include the vocabulary coverage rate ξ; the indicators for the accuracy and generation performance of the model include Top-n accuracy ($A_{\text{Top-N}}$), perplexity ψ, Mean Absolute Error (MAE), Precision, Recall, and F1 score.

The formula for the average path length λ is:

$$\lambda = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\sum_{j=1}^{M_i}\sum_{k=1}^{\mathcal{L}_{ij}} d(v_k, v_{k+1})}{M_i}\right) \tag{1}$$

In Formula (1), N depicts the total number of reasoning tasks; M_i depicts the number of possible reasonings for the i-th task; $\mathcal{L}_{ij}$ represents the specific length of the j-th path in the i-th task; $d(v_k, v_{k+1})$ represents the distance between node v_k and node v_{k+1}.

The indicator of semantic similarity is generally represented by the cosine similarity between two text word vectors. The formula for vocabulary coverage ξ is:

$$\xi = \frac{\sum_{i=1}^{N}\delta(w_i)}{\sum_{i=1}^{N}\sigma(w_i)} \times 100\% \tag{2}$$

Among them, $\delta(w_i)$ and $\sigma(w_i)$ are both indicator functions. When the word w_i is correctly recognized by the model as a professional term, $\delta(w_i) = 1$, otherwise it is 0; when the word w_i is a professional term, $\sigma(w_i) = 1$, otherwise it is 0.

Top-n accuracy $A_{\text{Top-n}}$ is the frequency of correct answers being included in the first n predictions, and the formula is:

$$A_{\text{Top-n}} = \frac{1}{T}\sum_{i=1}^{T}\left(\sum_{j}^{n}\delta(t_i \in t_{i,j})\right) \tag{3}$$

When the true label t_i is among the first n possible outcomes predicted by the model, $\delta(t_i \in t_{i,j}) = 1$, otherwise it is 0.

The perplexity ψ refers to the predictive ability of the language model for the next word, and the formula is:

$$\psi = 2^{-\frac{1}{T}}\sum_{i=1}^{T}\sum_{j=1}^{\varepsilon_i}\log_2 P(\eta_{i,j}|\eta_{i,<j}) \tag{4}$$

Where ε_i shows the number of words in the i-th sample; $\eta_{i,j}$ shows the j-th word in the i-th sample; $\eta_{i,<j}$ represents all words before the jth word in the i-th sample; $P(\eta_{i,j}|\eta_{i,<j})$ shows the probability of predicting the next word given a previous word sequence. MAE, Precision, Recall, and F1 values are indicators commonly utilized to determinemodel accuracy and generation performance, and the formulas are no longer written.

Table 1 Data collection status.

Document Type	Total Number of Documents	Number of Training Set Documents	Number of Validation Set Documents	Number of Test Set Documents
Maintenance Record of Power Equipment	400	280	60	60
Technical Tutorial Document for The Power Industry	500	350	75	75
Power Industry Policy Document	70	49	11	10
Standard Specification Documents for The Power Industry	600	420	90	90
Excellent Paper Documents in The Power Industry	65	45	10	10

Table 2 Schematic table of part of dataset preprocessing.

Document Number	Document Type	Partial Participle Content
0016	Maintenance Record of Power Equipment	transformer oil, cooling, aging, oil tank cleaning, …
0537	Technical Tutorial Document for the Power Industry	insulated gloves, insulated boots, half bridge switching power supply,…
0602	Power industry policy document	power grid, operation, electricity market, …
1109	Standard specification documents for the power industry	cable, instance, substation, …
1632	Excellent paper documents in the power industry	circuit breaker, low voltage incoming line, withstand voltage, …

4. SIMULATION EXPERIMENT

4.1 Experimental Setup and Data Collection

During the experimental preparation phase, a dataset containing textual data in the field of electricity wascollected. The dataset wassourced from Dianzhi Network. The collected dataset waspartitioned into training set, validation set and test set in the ratio of 70%, 15% and 15% respectively. The relevant dataset information is displayed in Table 1.

Table 1 shows the specific situation of the collected dataset and the data volume of each dataset after partitioning. It can be seen that the dataset comprises1635 data documents related to electricity knowledge from Dianzhi Network, covering different aspects of the electricity field.

The hardware configuration used for the experiment comprisedNVIDIA Tesla V100 GPU, Intel Xeon E5-2690 v4 CPU, and 128GB of memory; the software configuration includedUbuntu 20.04 operating system, PyTorch 1.7 deep learning framework, and Python 3.8 programming language; the device wasa GPU (Graphics Processing Unit) server. During the training process, the model usedthe Adam optimizer with an initial learning rate of 0.00005, a batch size of 32, 10 training epochs, and an early stopping strategy.

Table 2 shows some examples of the dataset after data cleaning and segmentation.

From Table 2, it can be seen that during data preprocessing, the collected data documents are numbered and classified accordingly. In addition, data cleaning, document labeling, word segmentation, stop word removal, and lowercase conversion are performed on the document content.

4.2 Experimental Results

The preprocessed training set and validation set data wereimported into the BERT model that integrates the power knowledge graph for model training and parameter tuning. The parameter configuration of the BERT model integrated with apower knowledge graph is displayed in Table 3.

Table 3 lists the initial configuration, optimal configuration, and parameter range of the BERT model that integrates power knowledge graph. It can be clearly learned from Table 3 that in the parameter tuning of the BERT model that integrates a power knowledge graph, the experiment focuses more on the model's learning rate, batch size, training rounds, hidden layer size, attention head number, weight decay, dropout probability, maximum sequence length, and knowledge embedding dimension. By adjusting the configuration of these parameters, the model can ensure optimal performance in NLU tasks after integrating the power knowledge graph. Additionally, it can be noted that the batch size of the model has been optimized from 32 to 16, reducing the number of samples updated each time; the number of iterations has been optimized from 10 to 15, increasing the number of training rounds by 5. The hidden layer size, number of attention heads, weight decay, and dropout probability of the model remain unchanged at 768, 12, 0.01, and 0.1. The maximum sequence length of the model has been optimized from 128 to 256, increasing the length of text paragraphs processed by the model. The knowledge embedding dimension of the model has been optimized from 100 to 200, improving its ability to represent complex relationships and concepts in the power knowledge graph.

Table 3 Parameter configuration of BERT model integrating power knowledge graph.

Parameter	Initial Configuration	Optimal Configuration	Parameter Range
Learning Rate	0.00005	0.00003	[0.00001, 0.00005]
Batch Size	32	16	[8, 32]
Epochs	10	15	[10, 30]
Hidden Size	768	768	[768, 1024]
Attention Heads	12	12	[8, 16]
Weight Decay	0.01	0.01	[0, 0.1]
Dropout Probability	0.1	0.1	[0.1, 0.3]
Max Sequence Length	128	256	[128, 512]
Knowledge Embedding Dimension	100	200	[50, 300]

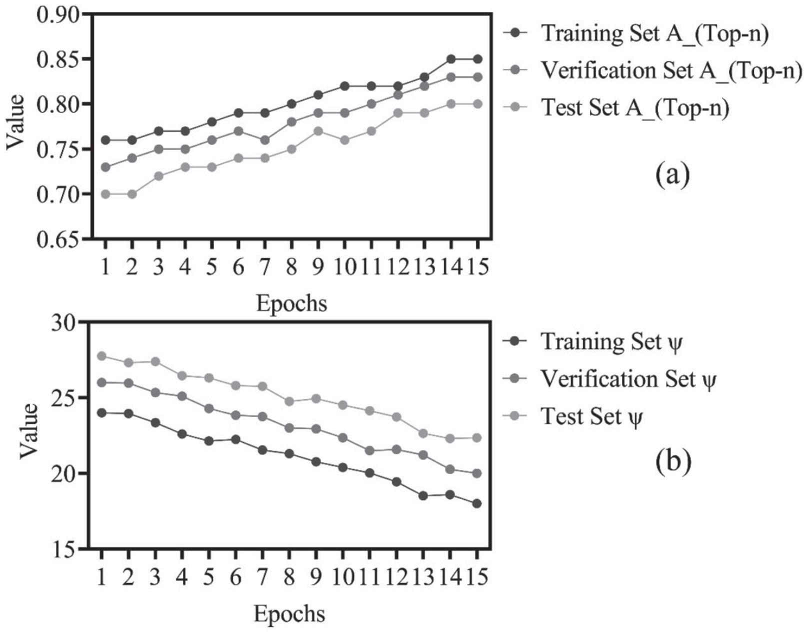

Figure 3 Performance of Top-n accuracy and perplexity in the training, validation, and test sets. Figure 3(a) Performance of Top-n accuracy on various datasets. Figure 3(b) Performance of confusion in various datasets.

Figure 3 shows the variation of Top-n accuracy and perplexity of the BERT model fused with power knowledge graph on the training set, validation set, and test set with the number of iterations.

Figure 3 illustrates the performance of Top-n accuracy and perplexity on the training, validation, and test sets. Top-n accuracy is utilized to assess the accuracy of model predictions, with a higher value indicating a higher proportion of correct answers in the first n results. From Figure 3(a), it can be learned that the Top-n accuracy of the BERT model that integrates the power knowledge graph increases with the number of iterations on all datasets. The perplexity level is utilized to assess the text quality generated by the model, with lower values showing that the generated text is closer to the real corpus. In Figure 3(b), the perplexity of the model decreases with increasing iteration times on all datasets. Furthermore, the perplexity of the training and test sets does not change significantly in the 13th to 15th iterations, indicating that the perplexity tends to stabilize and reaches a convergence state; the perplexity of the validation set does not change significantly during the 14th and 15th iterations. The Top-n accuracy tends to stabilize in all datasets at the 14th iteration.

Figure 4 shows the BERT model's performance incorporating the power knowledge graph for MAE, Precision, Recall, and F1 values.

Figure 4 shows the variation of MAE, Precision, Recall, and F1 values of the BERT model integrating power knowledge graph on the training set, validation set, and test set with the number of iterations. As shown in Figure 4(a), the MAE of the model gradually decreases with the increase of iteration times. MAE represents the mean absolute difference between the predicted value and the true value, and a lower MAE indicates a smaller prediction error of the model. Figures 4(b), (c), and (d) show that Precision, Recall, and F1 values fluctuate to some extent in the training, validation, and test sets, but show an increasing trendoverall. These indicators are commonly used to evaluate model performance, and the larger the value, the stronger the overall performance of the model. From the comprehensive performance depicted in Figures 4(b), (c), and (d), it is evident that the BERT model integrating the power knowledge graph performs well in terms of Precision, Recall, and F1 values.

Figures 3 and 4 show that the metrics in the training set outperform the validation set overall, while the validation set outperforms the test set, with only a few exceptions in a few

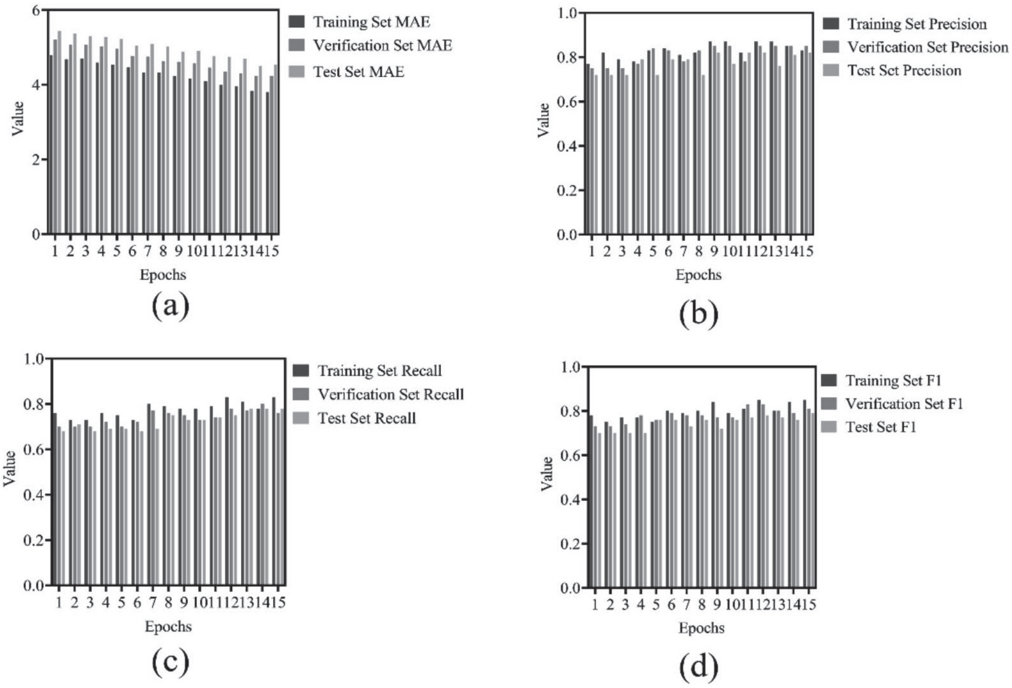

Figure 4 Performance of MAE, Precision, Recall, and F1 values on the training set, validation set, and test set. Figure4(a) Performance of MAE. Figure.4(b) Performance of Precision. Figure4(c).Performance of Recall. Figure4(d).Performance of F1 values.

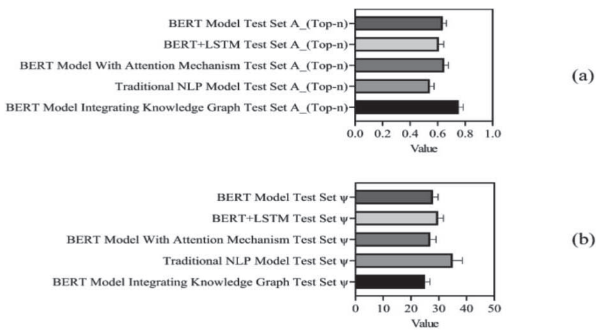

Figure 5 Comparison of Top-n accuracy and perplexity of various models. Figure 5(a). Comparison of Top-n accuracy among various models. Figure5(b). Comparison of perplexity among various models.

iterations. The reason for this phenomenon is that the model is trained on the training set and the parameters are adjusted on the validation set, so it performs better on both sets. It is normal for the model to perform slightly worse than the training and validation sets when faced with unknown data on the test set. From the data of the BERT model integrating power knowledge graph on the test set, although the test set does not perform as well as the training and validation sets, it still maintains at a high level, which indicates that the model can perform strongly.

4.3 Model Comparison

Figure 5 shows the comparison of Top-n accuracy and perplexity indicators between the BERT model fused with a power knowledge graph and the other four models on the test set.

Figure 5 demonstrates the comparison of Top-n accuracy and perplexity metrics for each model on the test set. From Figure 5 (a), it can be learned that the BERT model that integrates the power knowledge graph has a higher Top-n accuracy than the other four models, followed by the BERT model with an improved attention mechanism, whilethe traditional NLP model performs the worst. Figure 5 (b) shows that of the five models, the BERT model that integrates the power knowledge graph has the lowest perplexity, followed by the BERT model with an improved attention mechanism, and the traditional NLP model has the highest perplexity. Among them, the BERT model and the attention mechanism-improved BERT model have similar performance in Top-n accuracy and perplexity, while the traditional NLP model performs the worst in Top-n accuracy and ψ aspect.

Figure 6 shows the comparison of each model's MAE, Precision, Recall, and F1 values for the test set.

Figure 6 shows the performance of each model on the MAE, Precision, Recall, and the F1 values on the test set. A comparison of Figures 6(a), (b), (c), and (d) concludes that the BERT model incorporating the power knowledge graph is the best in the results of model comparison in terms of MAE, Precision, Recall, and F1 value. The values of both MAE, Precision, Recall and F1 are significantly better than

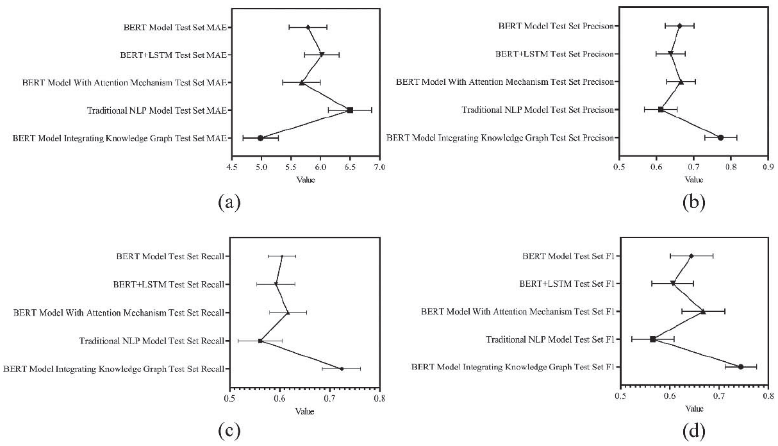

Figure 6 Comparison of MAE, Precision, Recall, and F1 values of various models on the test set. Figure 6 (a): Comparison of MAE. Figure 6 (b): Comparison of Precision. Figure 6(c). Comparison of Recall. Figure 6(d). Comparison of F1 values.

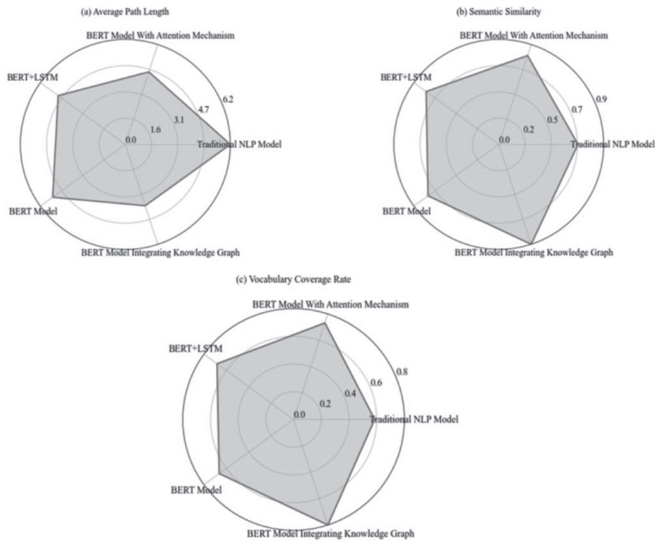

Figure 7 Performance evaluation metrics results foreach model in NLU tasks. Figure 7(a) Indicator results of the average path length λ of each model in the NLU tasks. Figure 7(b) Indicator results of semantic similarity of each model in NLU tasks. Figure 7(c). Indicator results of vocabulary coverage ξ for each model in NLU tasks.

those of the other four models. Except for the BERT model incorporating the power knowledge graph, the BERT model improved by the Attention mechanism performs the best and outperforms the other models in terms of MAE, Precision, Recall, and F1 value.

4.4 Performance Evaluation of BERT Model Integrating Power Knowledge Graph in NLU Tasks

Figure 7 shows the evaluation of the average path length λ, semantic similarity, and vocabulary coverage ξ of each model in the NLU task.

Figure 7 shows the radar chart of performance evaluation metrics for each model in the NLU task. Figure 7(a) shows the comparison results of each model in terms of average path length λ. The average path length λ is used to evaluate the path reasoning ability of each model; the smaller theλ, the higher is the reasoning ability. Figure 7(a) clearly shows that the traditional NLP model has the highest λ at 6.2; next is the BERT model, which is 5.3; the smallest λ (3.8) is the BERT model that integrates power knowledge graph. The semantic similarity between the text generated by each model and the reference text is calculated by calculating the similarity cosine, with the larger similarity indicating that the text generated by the model is closer to the reference text. Figure 7(b) shows that the semantic similarity of the BERT model fused with the power knowledge graph is the highest, at 0.9; next is the BERT model with improved attention mechanism, which is 0.8; the remaining models do not reach 0.8 in terms of semantic similarity. The vocabulary coverage ξ evaluates the ability of each model to process terminology pertaining to the field of electricity. It measures the ability of each model to process such terms by calculating the proportion of relevant terms that can be correctly identified and processed by each model when processing electricity-related text, and compare this to the total number of professional terms. The larger the vocabulary coverage ξ, the stronger the model's ability to handle professional terminology. It can be clearly seen from Figure 7(c) that the vocabulary coverage ξ of the BERT model that integrates the power knowledge graph is the highest at 0.8;

next, at 0.75is the BERT model with the improved attention mechanism; the traditional NLP model performs the worst, with a vocabulary coverage rate ξ of only 0.6.

5. CONCLUSIONS

In this study, the performance of BERT models in NLU tasks was enhanced by integrating power knowledge graphs. The problem of poor adaptability and the weak knowledge fusion ability of traditional language models whenhandling specific tasks in the field of electricity, which leads to the inability to fully capture deep relationships in context, was examined. Hence, a BERT model that integrated a power knowledge graph was constructed to solve this problem. The results of simulation experiments verified that the BERT model integrating a power knowledge graph performed well on NLU tasks in the power field. However, there are also shortcomings in this article. The construction and preprocessing process of the dataset is relatively complex and time-consuming. In addition, the exploration of other factors that may affect the performance of the model is not comprehensive. In future research, it is necessary to further optimize the fusion mechanism of knowledge graph and BERT model, and extend this method to other application fields in order to promote the development of NLU tasks in more fields.

FUNDING

State Grid Shanxi Electric Power Company Technology Project “Research on Intelligent AI Customer Service Technology for Supplier Service Center Based on Speech Recognition and Knowledge Graph Technology” (5205C0240004).

REFERENCES

1. Ozcift A, Akarsu K, Yumuk F, et al. Advancing natural language processing (NLP) applications of morphologically rich languages with bidirectional encoder representations from transformers (BERT): An empirical case study for Turkish. Automatika: èasopis za automatiku, mjerenje, elektroniku, raèunarstvoikomunikacije, 2021, 62(2): 226–1.
2. Feng X, Dai Y, Ji X, et al. Application of natural language processing in HAZOP reports. Process Safety and Environmental Protection, 2021, 155: 41–1.
3. Min B, Ross H, Sulem E, et al. Recent advances in natural language processing via large pre-trained language models: A survey. ACM Computing Surveys, 2023, 56(2): 1–1.
4. Bilal M, Almazroi A A. Effectiveness of fine-tuned BERT model in classification of helpful and unhelpful online customer reviews. Electronic Commerce Research, 2023, 23(4): 2737–1.
5. Qiao Y, Zhu X, Gong H. BERT-Kcr: prediction of lysine crotonylation sites by a transfer learning method with pre-trained BERT models. Bioinformatics, 2022, 38(3): 648–1.
6. Li M, Chen L, Zhao J, et al. Sentiment analysis of Chinese stock reviews based on BERT model. Applied Intelligence, 2021, 51: 5016–1.
7. Ma K, Tan Y, Tian M, et al. Extraction of temporal information from social media messages using the BERT model. Earth Science Informatics, 2022, 15(1): 573–1.
8. Chriqui A, Yahav I. HeBERT and HebEMO: a Hebrew BERT model and a tool for polarity analysis and emotion recognition. INFORMS Journal on Data Science, 2022, 1(1): 81–1.
9. Uveges I, Ring O. HunEmBERT: a fine-tuned BERT-model for classifying sentiment and emotion in political communication. IEEE Access, 2023, 11: 60267–1.
10. Eke C I, Norman A A, Shuib L. Context-based feature technique for sarcasm identification in benchmark datasets using deep learning and BERT model. IEEE Access, 2021, 9: 48501–1.
11. Geetha M P, Renuka D K. Improving the performance of aspect-based sentiment analysis using fine-tuned Bert Base Uncased model. International Journal of Intelligent Networks, 2021, 2: 64–1.
12. Alaparthi S, Mishra M. BERT: A sentiment analysis odyssey. Journal of Marketing Analytics, 2021, 9(2): 118–1.
13. Cai L, Song Y, Liu T, et al. A hybrid BERT model that incorporates label semantics via adjustive attention for multi-label text classification. IEEE Access, 2020, 8: 152183–1.
14. Sun Q, Xi X, Sun J, et al. Stance detection with a multi-target adversarial attention network. ACM Transactions on Asian and Low-Resource Language Information Processing, 2022, 22(2): 1–1.
15. Suresh Kumar K, Helen Sulochana C, Radhamani A S, et al. Sentiment lexicon for cross-domain adaptation with multi-domain dataset in Indian languages enhanced with BERT classification model. Journal of Intelligent & Fuzzy Systems, 2022, 43(5): 6433–1.
16. Song K, Tan X, Qin T, et al. Mpnet: Masked and permuted pre-training for language understanding. Advances in Neural Information Processing Systems, 2020, 33: 16857–1.
17. Wang B, Kuo C C J. Sbert-wk: A sentence embedding method by dissecting bert-based word models. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2020, 28: 2146–1.
18. Jiang Z H, Yu W, Zhou D, et al. Convbert: Improving BERT with span-based dynamic convolution. Advances in Neural Information Processing Systems, 2020, 33: 12837–1.
19. Yasunaga M, Bosselut A, Ren H, et al. Deep bidirectional language-knowledge graph pretraining. Advances in Neural Information Processing Systems, 2022, 35: 37309–1.
20. Ji S, Pan S, Cambria E, et al. A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(2): 494–1.
21. Koloski B, Perdih T S, Robnik-Sikonja M, et al. Knowledge graph informed fake news classification via heterogeneous representation ensembles. Neurocomputing, 2022, 496: 208–1.
22. Meng F, Yang S, Wang J, et al. Creating knowledge graph of electric power equipment faults based on BERT–BiLSTM–CRF model. Journal of Electrical Engineering & Technology, 2022, 17(4): 2507–1.
23. Qi Pengnian, Liao Yulun, and Qin Biao. A review of research on Chinese named entity recognition based on deep learning. Journal of Chinese Mini-Micro Computer Systems, 2023, 44 (09): 1857–1.
24. Lamba, V., Šimková, N., & Rossi, B. (2019). Recommendations for smart grid security risk management. Cyber-Physical Systems, 5(2), 92–118.
25. Shi Zhengyi, Lv Junke, Huang Hong. Research on the Construction of Urban Underground Pipeline Information Knowledge Graph Based on Neo4j. Journal of Safety Science and Technology, 2024, 20 (06): 5–1.

26. Xin Ji, Chengyue Yang, Xiaolong Zhao, Linxiao Dong, Yuliang Zhao. Application of data center knowledge graph based on powersystem fusion algorithm design. Engineering Intelligent Systems, 2023, 30(4): P 265–1.
27. Ningfang Wang. Personalized learning of college English using knowledge graphs combined with user portraits. Engineering Intelligent Systems, 2025, 33 (3): 339–1.
28. Niu Shiyin, Yu Junhe, Zhan Hongfei, et al. Research on Industrial Cluster Analysis Method Based on Graph Embedding Technology. Journal of Ningbo University (Natural Science & Engineering Edition), 2023, 36 (05): 29–1.
29. Zhao, S., & Wu, Z. (2021). Optimisation model of micro grid dispatching for smart building cluster based on blockchain. Cyber-Physical Systems, 8(2), 138–171.
30. Huang Dongling, Liu Yuan, Yuan Xiaoshuai, et al. Research on Data Mining of China's Hydrogen Energy Policy Based on Natural Language Processing Technology. Bulletin of Chinese Academy of Sciences, 2024,39 (06): 1032–1.
31. Rong Rong. Research on the Application of Natural Language Processing Technology-Assisted Writing Based on Artificial Intelligence. Science and Technology Innovation, 2024, (11): 96–1.
32. Mao Mingyi, Wu Chen, Zhong Yixin, et al. BERT Named Entity Recognition Model with Self Attention Mechanism. CAAI Transactions on Intelligent Systems, 2020, 15 (4): 772–1.
33. Wang Wenjuan, He Xiaolian, Hu Feng, et al. Multi-attention mechanism BERT classification model and its application in power operation and maintenance. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2023, 35 (01): 156–1.
34. LvXueqiang, Tian Chi, Zhang Le, et al. A multimodal sentiment analysis model that integrates multiple features and attention mechanisms. Data Analysis and Knowledge Discovery, 2024, 8 (05): 91–1.