A BERT-Based Question Answering System for Optical Fiber Transmission Networks

Haiyang Wang[1], Yanhui Lu[1], Yunxiao Wang[1], Chi Zhang[1], Ming Huang[2],

Mingqiang Zhu[2]

[1]Hebei Jixiangtong Electronic Technology Co., China

[2]Electronic Information Engineering College, Beijing Jiaotong University, China

Received: March 20, 2026; Revised: April 23, 2026; Accepted: May 25, 2026; Published: June 03, 2026

Abstract

To help operations staff quickly get technical support and decision-making basis for highway fiber-optic transmission networks, this paper plans to develop a BERT-based Question Answering (QA) system. The system first builds a knowledge graph to store related knowledge. Then, it uses the BERT model to identify entities and intents in user questions, matches them in the knowledge graph via template matching, and returns answers. This system offers efficient information query and technical support for highway fiber-optic transmission networks. Experimental results show that BERT model has 10.01%, 6.21%, 3.72% higher F1-score than TextRNN, TextCNN, Fasttext in intent recognition. BERT-BiLSTM-CRF model has 10.01%, 6.21%, 3.72% higher F1-score than LSTM, BiLSTM, BiLSTM-CRF in entity recognition.

Keywords: BERT, Question Answering System, Knowledge Graph, Named Entity Recognition, Intent Recognition.

1 Introduction

China's expressway communication network has evolved into a nationwide fiber-optic infrastructure. However, traditional manual fault diagnosis suffers from delayed responses and low knowledge reusability amid massive operation and maintenance (O&M) data. While natural language processing offers new intelligent solutions, including rule-based question-answering (QA) systems [1], current applications face three major challenges. First, there is insufficient semantic understanding in specialized domains, with traditional Named Entity Recognition (NER) F1-score generally remaining below 75%. Second, limited knowledge-based dimensions fail to accurately model dynamic network associations. Third, there is a distinct lack of reliable assurance mechanisms for critical infrastructure decisions.

To address these gaps, previous studies have explored Bi-directional Long Short-Term Memory (BiLSTM)-Conditional Random Field (CRF) for entity extraction and Neo4j-based knowledge graphs, though they still face long-range dependency and response time bottlenecks. Additionally, while Trusted Platform Control Module (TPCM) technology provides trusted computing in cloud environments [2], its integration with QA systems remains unreported. Building upon foundational data in Optical Transport Network (OTN) technology [3], this paper props an intelligent Bidirectional Encoder Representations from Transformers (BERT)-based QA system that integrates deep semantic understanding, dynamic knowledge reasoning, and trusted assurance to enable precise O&M of fiber-optic networks.

2 System design methodology

This chapter focuses on the architecture and implementation of an intelligent QA system designed for optical fiber transmission networks. The core modules of the proposed system consist of a BERT-based entity and intent recognition module, alongside a knowledge graph-based answer retrieval module. This chapter comprehensively analyzes the underlying principles of the aforementioned models, detailing the construction of the knowledge graph [4] as well as the knowledge graph-based query methodology utilizing template matching [5].

2.1 System Workflow

By leveraging structured semantic associations [6], knowledge graph technology establishes a novel knowledge representation paradigm for the intelligent O&M of optical fiber networks by utilizing an entity-relation-attribute three-dimensional modeling framework. In contrast to traditional relational databases constrained by two-dimensional table structures, the proposed knowledge graph facilitates the dynamic chain mapping of Optical Distribution Network topologies. Furthermore, it enables multi-level causal reasoning that integrates fault knowledge with standard parameters and conducts time-series analysis of fiber performance degradation based on timestamp attributes.

This mechanism significantly reduces query response latency while effectively overcoming inherent bottlenecks in traditional O&M systems such as poor real-time performance in topology reasoning and unidimensional data associations. The overall system workflow proceeds by parsing the user input query to extract entity and intent features before executing the retrieval within the knowledge graph and ultimately returning the target answer.

2.2 BERT-based question comprehension

For Chinese natural language processing tasks, Cui et al. [7] proposed a whole-word mask (WWM) pre-training strategy. Inspired by this, this paper utilizes the BERT model for intent recognition to fully leverage its deep semantic modeling capabilities in the Chinese context. The Transformer architecture abandons traditional recurrent structures in favor of a self-attention mechanism, significantly improving the model's parallelization and training efficiency. At the input representation level, BERT accurately captures deep semantic and positional features through the fusion of character, sentence, and position embeddings.

During the pre-training stage, BERT employs a masked language model (MLM) to randomly mask 15% of the vocabulary in the input sequence, forcing the model to utilize bilateral predictions. After pre-training, the model can efficiently adapt to downstream tasks through specific fine-tuning. In vertical fields such as fiber optic transmission networks, BERT can extract and parse alarm information, automatically generate Cypher query templates, and significantly improve the response efficiency of intelligent O&M.

For the extraction of intent information, a [CLS] token and a [SEP] token are respectively appended to the beginning and end of the user query sequence. The BERT model processes this sequence to obtain contextual semantic representations. For intent classification, the hidden state h_0 corresponding to the [CLS] token is directly passed into a fully connected layer and a SoftMax activation function to calculate the probability distribution of the intent labels.

$$P = Softmax(h_0 W^o + b^o) \tag{1}$$

In this expression, W_o represents the weight matrix while b_o denotes the bias term and K signifies the total number of classification labels which determines the dimensionality of the output probability vector P.

LSTM networks have demonstrated significant efficacy in capturing long-range dependencies when processing sequence data. To enhance entity extraction accuracy, this study constructs a BERT-BiLSTM-CRF model utilizing the Begin, Inside, Outside (BIO) tagging scheme. The input sequence is first transformed by

the BERT layer into a feature vector encapsulating rich contextual semantics. This vector is then fed into the BiLSTM layer, which captures global sequential features by computing and concatenating the forward $\overrightarrow{h_t}$ and backward $\overleftarrow{h_t}$ hidden states:

$$h_t = < \overrightarrow{h_t}, \overleftarrow{h_t} > \tag{2}$$

The concatenated contextual information h_t is subsequently processed through a hyperbolic tangent activation function to output O_t.

Following the BiLSTM layer, a CRF layer is introduced to explicitly model the direct dependencies between adjacent labels. Let $X = (x_1, x_2, \cdots, x_n)$ denote the input sequence to the CRF layer, and $Y = (y_1, y_2, \cdots, y_n)$ denote the predicted entity label sequence. Employing maximum likelihood estimation, the CRF layer calculates the optimal probability of the predicted sequence. The score computation and the final probability output are formulated as follows:

$$score(X, Y) = \sum_{i=1}^{n} \left(P_{i,y_i} + W_{y_i, y_{i+1}} \right) \tag{3}$$

$$P(Y|X) = \frac{exp(score(x,y))}{\sum_{y'} exp(score(x,y'))} \tag{4}$$

where $P_{i,j}$ denotes the score of assigning the j-th label to the i-th word, and $W_{i,j}$ represents the transition score from label i to label j.

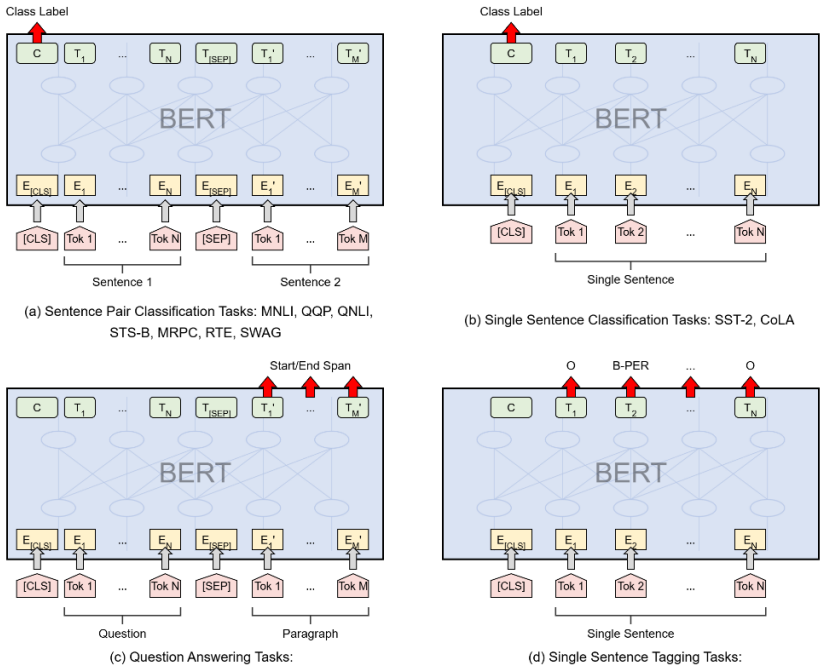


Figure 1. BERT Fine-Tuning Diagram

2.3 Answer queries based on knowledge graphs

The core of constructing a knowledge graph for optical fiber transmission networks lies in the acquisition and standardization of multi-source data. By collecting professional web pages and scientific literature from the

National Science and Technology Library [8], a solid foundation is established for defining graph entities and relationships. The underlying data model utilizes an entity-relationship-entity triplet format that maps entities as nodes and relationships as directed edges to capture deep semantic associations and support complex fault reasoning efficiently.

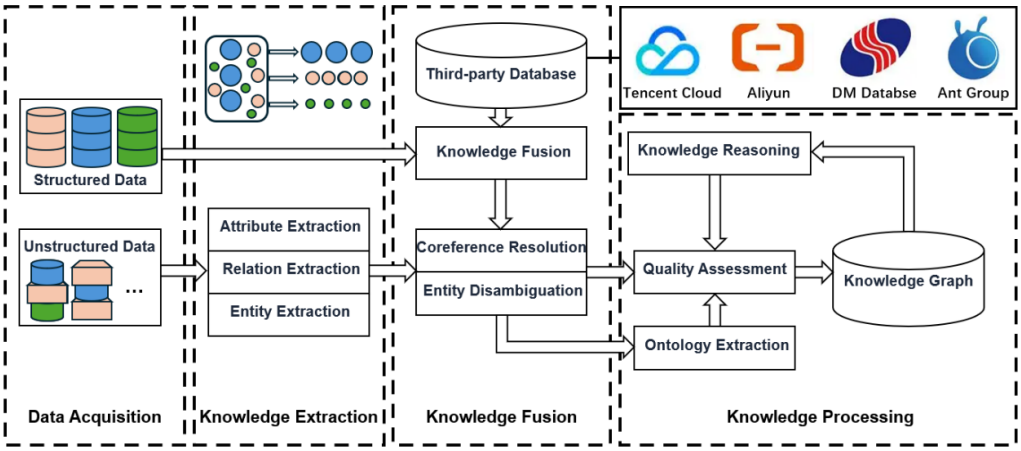


Figure 2. Knowledge Graph Construction Architecture

Since raw data contains substantial unstructured text, natural language processing techniques are utilized for initial knowledge extraction [9]. These extracted triplets are integrated with structured business data for entity alignment and reasoning analysis, uncovering hidden topological or causal relationships to form a high-quality domain map. The defined schema constitutes the ontological specification, outlining five core entity types, including network nodes, optical fibers, network topology, network services, and performance indicators, as well as four semantic relationships, namely connection, inclusion, provision, and impact. Following these constraints, the fused data is instantiated and stored persistently within a Neo4j graph database.

During the query stage, the system parses user questions through the intent and entity recognition module to extract key parameters. A template matching method [10] is then deployed to map the extracted entities and intents directly to specific query paths. Based on these mapped parameters, the system automatically generates Cypher statements to execute multi-hop retrievals in the Neo4j database, structuring the hit results to return a precise solution to the user.

2.4 Reliability requirements for transmission QA system

Applying intelligent question answering systems to critical information infrastructure, such as highway optical fiber networks, demands stringent credibility. The proposed system satisfies five key security and compliance dimensions.

To ensure the rigorous protection of sensitive critical information infrastructure (CII) data, the system mandates strict adherence to the Data Security Law [11] and the Cross-border Data Transfer Measures [12]. Beyond basic privacy compliance, the architecture incorporates federated learning to neutralize potential gradient leakage in natural language processing models, ensuring that all training operations remain consistent with the data minimization and encrypted storage guidelines established in the Personal Information Security Specification [13].

The security posture of the system is further fortified by requiring operators to resist sophisticated network attacks [14], a necessity given that knowledge graph integrity directly dictates the reliability of operation and maintenance. To this end, the system employs a dynamic trusted verification mechanism. This specific layer of defense ensures that the hash fingerprints of the graph data consistently meet the Level 3 cryptographic benchmarks set by GB/T 39786-2021 [15].

Service reliability remains a paramount concern for transportation intelligent systems, which demand a minimum of 99.99% availability [16]. To eliminate single points of failure, the architecture utilizes active-active redundancy a configuration rigorously verified through Markov chain modeling to fulfill ISO/IEC 27001 business continuity requirements.

Finally, addressing the black box nature of deep learning is critical as all high-risk artificial intelligence decisions within the system must be rendered traceable [17]. The system achieves this by generating explicit causal reasoning chains while the use of knowledge graph entity relationships ensures compliance with the ITU Trusted AI Evaluation Framework [18]. To fulfill long-term compliance audit mandates, the system retains operation logs for at least six months [19]. By utilizing blockchain technology for full-link certificate storage during federated learning, the system achieves the fine-grained traceability required by NIST SP 800-171 Rev.3 to ensure every decision remains verifiable and tamper-proof.

3 Distributed training modeling based on trusted federated learning

To address the joint modeling needs of multiple expressway management units, this paper proposes a distributed collaborative training framework utilizing privacy protection and trusted computing. It builds joint optimization objectives by combining weighted loss functions and regularization constraints from each participant's local data. The system features a two-layer security architecture comprising a Trusted Execution Environment for hardware-level isolation and a national encryption protocol designed to secure parameter interactions through ciphertext gradient aggregation.

The scheme details key training processes, including trusted verification, local gradient processing, and encrypted parameter updates, using formal methods to prove theoretical boundaries for privacy and convergence. It also integrates knowledge graph technology for cross-domain entity relationship mining and joint reasoning. Results confirm this architecture synergistically enhances training efficiency and knowledge discovery while ensuring high-level privacy.

The underlying mechanism draws inspiration from collaborative optimization in cutting-edge network systems. Similar concepts have successfully improved component reliability via intelligent co-balancing in cyber-physical systems [20], enhanced resource allocation flexibility through adaptive service configuration in network digital twins [21], and explored highly reliable architectures in decentralized networks. Ultimately, this trusted federated learning method aims to break business data silos, strictly balancing data privacy with global model performance across complex multi-agency environments.

3.1 Formalization of collaborative training problems

Assume there are M highway management entities (participants), each holding a private dataset $D_m = \left\{\left(x_i^{(m)}, y_i^{(m)}\right)\right\}_{i=1}^{N_m}$. Here, $x_i^{(m)} \in R^L$ denotes a sequence of word embeddings of length L for an input query, and $y_i^{(m)} \in Y$ represents the corresponding intent category label or the BIO tagging sequence for entities. The objective of federated learning is to determine the globally optimal BERT parameters θ^*, such that:

$$\theta^* = \underset{\theta}{argmin} \sum_{m=1}^{M} \frac{N_m}{N} \left[\frac{1}{N_m} \sum_{i=1}^{N_m} l\left(f_\theta\left(x_i^{(m)}\right), y_i^{(m)}\right) + \lambda R(\theta)\right] \quad (5)$$

Where $N = \sum_{m=1}^{M} N_m$, $l(\cdot)$ denotes the cross-entropy loss function, $R(\theta) = \| \theta \|_2^2$ represents the L2 regularization term, and λ is the trade-off coefficient. This objective function is derived from the SoftMax classification structure presented in this study, but has been extended to accommodate a multi-party setting.

3.2 Trusted Computing Component Definition

Each participant deploys a trusted container T_m based on Intel Trust Domain Extensions (TDX) technology. This container satisfies strict security properties. Specifically, for any given dataset D_m, the mutual information between the model parameters θ and the dataset under the condition of the given trusted container must be less than or equal to the privacy leakage upper bound ϵ_{priv}, expressed as $I(\theta; D_m|T_m) \leq \epsilon_{priv}$. Here, the mutual information is represented by the variable I, and the value of this privacy leakage upper bound is strictly less than 2^{-64}.

In the hierarchical homomorphic encryption mechanism, the system adopts the national cryptographic SM9 algorithm to implement parameter encryption. The specific operation involves transforming the local model gradient ∇L_m into a ciphertext format encrypted by the public key $[\![\nabla L_m]\!]_{pk}$. This encryption scheme inherently complies with the computational requirements of additive homomorphism and scalar multiplication.

3.3 Distributed training algorithm

In the distributed collaborative training framework, achieving efficient and secure multi-party joint modeling remains a crucial challenge. Inspired by intelligent collaborative tracking in the power Internet of Things and collaborative automation in multipath industrial networks, the trusted federated learning model proposed in this study aims to strictly balance data privacy protection with model performance improvements across multiple management units.

The process begins with the coordinating server generating the initial global BERT model, after which each trusted container verifies the model signature through a secure protocol.

During any given training round, the trusted container executes a series of sequential operations. Initially, it performs parameter downloading by decrypting the global model parameters using its local secret key.

Subsequently, the local gradient is calculated based on downloaded parameters:

$$\nabla L_m = \frac{\partial}{\partial \theta}\left[\frac{1}{N_m}\sum_{i=1}^{N_m} l\left(f_\theta\left(x_i^{(m)}\right), y_i^{(m)}\right) + \lambda \parallel \theta \parallel_2^2\right]\Big|_{\theta=\theta_g^{(t)}} \tag{6}$$

To ensure training stability, gradient clipping is applied. The system scales down the local gradient proportionally if its magnitude exceeds a predefined threshold, preventing excessive parameter updates.

Following this clipping procedure, the gradient is encrypted using the public key and securely uploaded back to the central server.

Finally, secure aggregation takes place at the server level. The server computes the global encrypted gradient by calculating the weighted average of the received local gradients according to their respective data sample sizes. Upon decryption using the SM9 private key, the server updates the global model parameters employing a standard gradient descent approach.

3.4 Formal security proof

Regarding the data privacy guarantee, under the semi-honest model and assuming the SM9 algorithm achieves IND-CPA security, the system ensures strong protection for any participant D_m. The probability that an adversary A successfully infers the original text $x_i^{(m)}$ from the encrypted gradients satisfies a specific upper bound:

$$Pr\left[A\left(\{[\![\nabla L_m]\!]_{pk}\}_{t=1}^{T}\right) \rightarrow x_i^{(m)}\right] \leq \frac{1}{|X|} + negl(\lambda) \tag{7}$$

Here $|X|$ denotes the vocabulary size, and λ is the security parameter. Furthermore, the security framework

can be formally demonstrated by constructing a simulator. For the established view of the adversary $V = \{[\![\nabla L_m]\!]_{pk}, \pi_m\}$, there exists a simulator S capable of generating an indistinguishable ciphertext $[\![\tilde{V}]\!]_{pk}$. Based on the semantic security inherent to the SM9 algorithm, the adversary view and the simulator output are computationally indistinguishable, expressed as $V \overset{c}{\approx} S$. Given the specific gradient expression of the BERT model $\nabla L_m = \sum h_i W^o$, the adversary remains fundamentally unable to reconstruct the input word embeddings corresponding to h_i.

3.5 Convergence theory analysis

Suppose the loss function satisfies:

L-smoothness:

$$\| \nabla L(\theta_1) - \nabla L(\theta_2) \|_2 \leq L \| \theta_1 - \theta_2 \|_2 \tag{8}$$

μ-Strong convexity:

$$L(\theta_1) \geq L(\theta_2) + \nabla L(\theta_2)^T(\theta_1 - \theta_2) + \frac{\mu}{2} \| \theta_1 - \theta_2 \|_2^2 \tag{9}$$

After T rounds of training, the upper bound of the global model error is:

$$E\left[\| \theta^{(T)} - \theta^* \|_2^2\right] \leq \left(1 - \frac{\mu}{L}\right)^T \| \theta^{(0)} - \theta^* \|_2^2 + \frac{4G^2}{\mu^2} \sum_{m=1}^{M} \left(\frac{N_m}{N}\right)^2 \sigma_m^2 \tag{10}$$

Where G denotes the upper bound of the gradient, and σ_m^2 represents the data variance of client m. As the data distribution among participants becomes similar, i.e., $\sigma_m^2 \to 0$, the convergence rate approaches that of centralized training.

3.6 Collaborative optimization with knowledge graph

The federated BERT model is comprehensively applied to the construction of the knowledge graph to enhance its overall capability.

In terms of entity extraction enhancement, the globally optimized model f_{θ^*} is utilized to extract novel triplets, such as the relationship between optical fiber equipment and specific fault types, directly from the unstructured logs of each participating entity.

Regarding cross-domain relational reasoning, the framework establishes joint reasoning rules through the central coordination server. For example, if a primary entity connects to a secondary entity, and this secondary entity subsequently impacts a tertiary entity, the system logically infers an indirect fault relationship between the primary and tertiary entities.

In the context of trusted storage, these newly discovered triplets are persistently written to the Neo4j database of each participant. This writing process occurs strictly after passing verification by the Trusted Execution Environment, ensuring the integrity and security of the distributed knowledge base.

4 Dual-system trusted recovery mechanism based on TPCM

The knowledge graph is the core component of the BERT-based QA system, directly impacting overall performance. To maintain high availability during service anomalies, this chapter proposes a TPCM-based dual-system trusted recovery mechanism. It utilizes dual-architecture protection components to automatically switch to a backup database while employing TPCM's dynamic integrity measurement to ensure recovery

credibility and data integrity.

The proposed TPCM-based trusted recovery mechanism elevates single-node system availability from 99.9% to 99.99% using a dual-system switching strategy and dynamic measurement. This ensures recovery verifiability and offers high engineering feasibility, integrating seamlessly into the existing QA system without altering the core BERT structure.

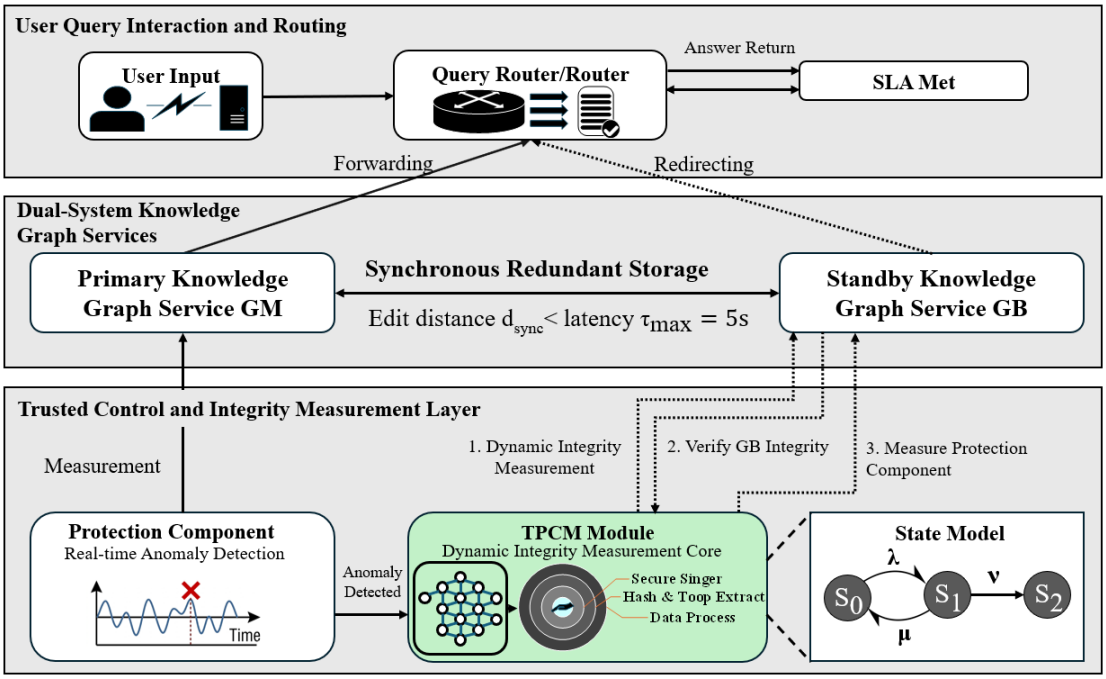


Figure 3. Dual-System Trusted Recovery Mechanism Based on TPCM

4.1 System architecture and state model

The system employs a dual-system architecture comprising a primary G_M and a standby G_B knowledge graph service. A protection component monitors the primary service in real-time, triggering a failover upon detecting anomalies. Concurrently, the TPCM module dynamically verifies the integrity of both the protection component and the standby database to ensure a trustworthy failover process.

Define the knowledge graph service state as Markov chains, $S = \{S_0, S_1, S_2\}$

S_0: The main knowledge graph G_M is running normally.

S_1: Anomaly detection status (TPCM metric failed).

S_2: Backup knowledge graph G_B takeover service.

The state transition probability matrix P satisfies:

$$P = \begin{pmatrix} 1-\lambda\Delta t & \lambda\Delta t & 0 \\ \mu\Delta t & 1-(\mu+\nu)\Delta t & \nu\Delta t \\ 0 & 0 & 1 \end{pmatrix} \quad (5)$$

Where λ denotes the failure rate of the primary repository (events/hour), μ denotes the repair rate, ν denotes the switchover failure rate, and Δt denotes the state detection interval, where $\Delta t = 1s$.

4.2 Dynamic integrity measurement and switching control

The integrity measurement value H_t of the knowledge graph $\mathcal{G}$ computed by the TPCM at time t is given by:

$$H_t = TPCM.Hash(G_t) \oplus TPCM.Sign_{sk}\big(t \parallel topo(G_t)\big) \tag{6}$$

Wherein, the fingerprint of the topology of the graph is represented, and the calculation method is:

$$topo(G) = \sum_{e \in V} SM3(e.type) + \sum_{r \in E} SM3(r.type) \tag{7}$$

The switching control algorithm is divided into the following steps.

- Synchronous redundant storage

Regarding synchronous redundant storage, the active and standby knowledge graphs must maintain synchronous storage. This ensures that the edit distance d_{sync} remains within a maximum configurable latency τ_{max}, which is set to five seconds in this system.

- Anomaly detection

TPCM periodically calculates measures $H_t^{(M)}$ and triggers switching when k consecutive detections fail:

$$\sum_{i=1}^{k} I\left(H_{t-i\Delta t}^{(M)} \neq H_{golden}^{(M)}\right) \geq k_{th} \tag{8}$$

$k = 3$, $k_{th} = 2$, to achieve a false positive rate $P_{false} < 0.1\%$.

- Trusted Switching

The system first verifies the integrity of the backup graph using the TPCM signature. Once verified, it performs an atomic switch, assigning the backup graph G_B to take over active services if the primary is not in the normal state S_0.

- Service recovery

The question answering system seamlessly redirects all incoming user queries to the newly active graph. The overall failover process is strictly bounded, adding a maximum delay, which is under 50 milliseconds, to the system's baseline response time of 200 milliseconds.

4.3 Usability analysis and mathematical proof

System steady-state availability A is calculated as:

$$A = \frac{MTTF}{MTTF+MTTR} \tag{9}$$

Among them, $MTTF = \frac{1}{\lambda}$, $MTTR = \frac{1}{\mu} + t_{switch}$.

With dual architecture:

$$MTTF_{dual} = \frac{1}{\lambda^2 \cdot t_{switch}} \tag{10}$$

By incorporating the empirically measured operational parameters, including a primary repository failure rate of $\lambda = 0.01$ events per hour and a system switchover time of $t_{switch} = 0.1s$ seconds, into the steady-state availability model for dual architectures, the calculation yields the following result:

$$A = \frac{1}{1+0.01\times0.1/3600} \approx 99.9996\% \quad (11)$$

SLA requirements met $A \geq 99.99\%$.

4.4 Integration with existing systems

To seamlessly integrate with existing systems, a TPCM verification node is inserted into the data processing pipeline immediately before writing data to the Neo4j database, ensuring data integrity at the source. During the query phase, the system router directly forwards incoming user queries to the currently active knowledge graph G_{active}. Furthermore, the backup graph G_B is optimized with memory caching so that its query time strictly does not exceed 1.2 times that of the primary graph G_M.

5 Analysis of results

In this chapter, the evaluation index of the evaluation model is first proposed, and then the intent recognition model and entity recognition model used in this paper are experimented and the performance of the model is evaluated, and finally the performance of the model is compared with other models to analyze the performance of the model.

5.1 Evaluation indicators

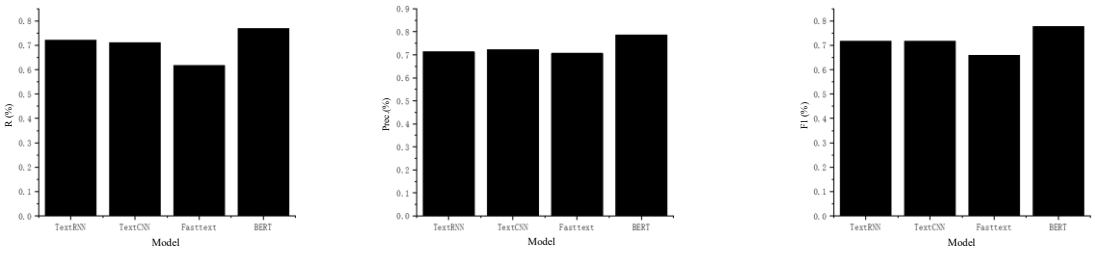
The main indicators evaluated by the experiment include precision (P), recall (R), and F1-score (F1), which collectively measure the model's performance in the intent recognition task.

5.2 Experimental results and analysis of results

In this paper, the BERT Base pre-trained model is used in the intent recognition task, which has 12 layers of encoders, 768 hidden units, and the number of attention mechanism heads is 12, and the maximum sequence length set in this paper is 40, and the learning rate is 1e-5. Table 1 shows the test results of each model, and Figure 4 shows the results of P, R, F1 value comparison.

Table 1. Intent Recognition Experimental Results Comparison

Model	*P*	*R*	*F*1
TextRNN	71.37%	72.19%	71.78%
TextCNN	72.31%	71.14%	71.72%
Fasttext	70.65%	61.87%	65.96%
BERT	78.74%	76.89%	77.80%



(a) P Value Comparison Chart (b) R Value Comparison Chart (c) F1 Value Comparison Chart

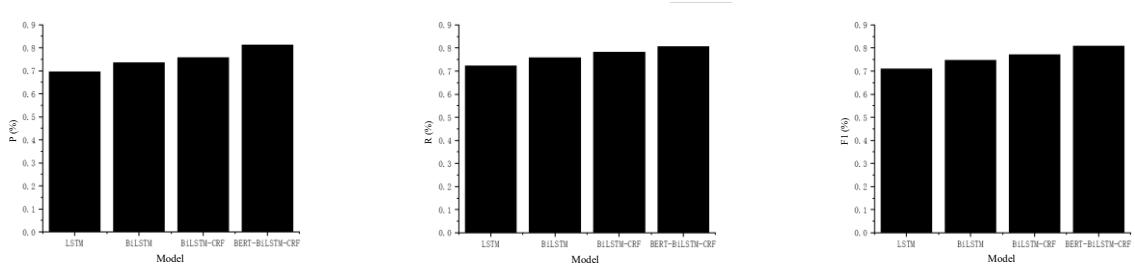
Figure 4. Intent Recognition Experiment Models' P, R, F1 Value Comparison Chart

The experimental results show that compared with other models, the accuracy, recall and F1 values of the BERT model in the intent recognition task of the medical QA system are 6.02%, 6.08% and 11.84% higher than those of TextRNN, TextCNN and Fasttext models, respectively This fully proves the excellent ability of the BERT model in capturing and understanding user semantics, and can efficiently and accurately complete the intent recognition task in the question and answer system.

In the entity recognition task, the BERT-BiLSTM-CRF model is used to carry out entity recognition experiments on the constructed dataset, in which the sequence length is set to 128, the hidden layer size of BiLSTM is 128, the learning rate is 1e-5, and it is compared with LSTM, BiLSTM and BiLSTM-CRF models. The experimental structures of the same dataset were compared. Table 2 **Error! Reference source not found.**shows the experimental results of each model, and**Error! Reference source not found.** Figure 5 shows the comparison results of P, R, and F1 values obtained by the entity recognition experiments of each model.

Table 2 NER Experimental Results Comparison

Model	P	R	F1
LSTM	69.58%	72.28%	70.90%
BiLSTM	73.62%	75.81%	74.70%
BiLSTM-CRF	75.84%	78.26%	77.19%
BERT-BiLSTM-CRF	81.23%	80.59%	80.91%



(a) P Value Comparison Chart (b) R Value Comparison Chart (c) F1 Value Comparison Chart

Figure 5. NER Experiment Models P, R, F1 Value Comparison Chart

By comparing the experimental data, it is obtained that the F1 value of BiLSTM is increased by 3.8% compared with the LSTM model, indicating that the semantic information obtained by BiLSTM has a better effect. The BERT-BiLSTM-CRF model used in this paper has 10.01%, 6.21% and 3.72% improvement in F1 value compared with LSTM, BiLSTM and BiLSTM-CRF models in the task of entity recognition. This paper shows that the model used in this paper can better complete the entity recognition task of the question answering system than other models.

6 Summary

This paper proposes a BERT-based QA system for optical fiber transmission networks, utilizing a constructed knowledge graph to deliver precise answers. By leveraging BERT's semantic capabilities, the system significantly improves user query understanding. However, current limitations include a small training dataset and a reliance on template matching, restricting responses to simple queries. Future improvements will focus on training with larger Chinese datasets and developing advanced retrieval mechanisms for complex user questions.

Acknowledgement

This work was supported by Hebei Jixiangtong Electronic Technology Co., Ltd. Science and Technology Project (No. 100115-07-2024-0022), the Fundamental Research Funds for the Central Universities (No. 2024YJS122), and the National Key Research and Development Program of China (No. 2023YFB2904200).

References

[1] Pan S, Luo L, Wang Y, et al. Unifying large language models and knowledge graphs: A roadmap[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(7): 3580-3599.
[2] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). 2019: 4171-4186.
[3] Guo X, Li Y, Zhang Z, et al. Application analysis of based on 100GB/s OTN technology in ultra long distance power communication network[C]//2021 IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). IEEE, 2021, 5: 1636-1639.
[4] Hofer M, Obraczka D, Saeedi A, et al. Construction of knowledge graphs: Current state and challenges[J]. Information, 2024, 15(8): 509.
[5] Abujabal A, Yahya M, Riedewald M, et al. Automated template generation for question answering over knowledge graphs[C]//Proceedings of the 26th international conference on world wide web. 2017: 1191-1200.
[6] d'Amato C, Masella P, Fanizzi N. An approach based on semantic similarity to explaining link predictions on knowledge graphs[C]//IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. 2021: 170-177.
[7] Cui Y, Che W, Liu T, et al. Pre-training with whole word masking for chinese bert[J]. IEEE/ACM transactions on audio, speech, and language processing, 2021, 29: 3504-3514.
[8] Song F, Ma Y, You I, et al. Smart collaborative evolvement for virtual group creation in customized industrial IoT[J]. IEEE Transactions on Network Science and Engineering, 2022, 10(5): 2514-2524.
[9] Uttam K. Extracting and Utilizing templates for Question Answering over Knowledge Graph[J]. 2021.
[10] Standing Committee of the National People's Congress. Data Security Law of the People's Republic of China [Z]. Beijing: China Legal Publishing House, 2021.
[11] Cyberspace Administration of China. Measures for the Security Assessment of Outbound Data Transfers [Z]. Beijing: Cyberspace Administration of China, 2022.
[12] Standardization Administration of China. Information security technology—Personal information security specification: GB/T 35273-2020 [S]. Beijing: Standardization Administration of China, 2020.
[13] State Council of the People's Republic of China. Regulations on the Protection of the Security of Critical Information Infrastructure [Z]. Beijing: State Council of the People's Republic of China, 2021.
[14] Standardization Administration of China. Information security technology—Baseline for information system cryptography application: GB/T 39786-2021 [S]. Beijing: Standardization Administration of China, 2021.
[15] International Telecommunication Union. Recommendation ITU-T G.1051: Interactivity test [S]. Geneva: International Telecommunication Union, 2023.
[16] UNESCO. Recommendation on the ethics of artificial intelligence [Z]. Paris: UNESCO, 2022.
[17] European Parliament and Council. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [Z]. Official Journal of the European Union, 2024.
[18] European Parliament and Council. Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union (NIS 2 Directive) [Z]. Official Journal of the European Union, 2022.
[19] Song F, Ai Z, Zhang H, et al. Smart collaborative balancing for dependable network components in cyber-physical systems[J]. IEEE Transactions on Industrial Informatics, 2020, 17(10): 6916-6924.
[20] Ma Y, Song F, Pau G, et al. Adaptive service provisioning for dynamic resource allocation in network digital twin[J]. IEEE Network, 2023, 38(1): 61-68.
[21] Song F, Ma Y, Yuan Z, et al. Exploring reliable decentralized networks with smart collaborative theory[J]. IEEE Communications Magazine, 2023, 61(8): 44-50.