ingegneria sismica

International Journal of Earthquake Engineering

Anno XLIII - Num. 2 - 2026

Construction of Digital Resource Map and Intelligent Retrieval Algorithm for Narrative of Yangtze River Culture

Yu Fang[1] *and Lili Zhang*[1,*]

[1] Jingchu University of Technology, Jingmen City, Hubei Province, China, 448000

SUMMARY: *As one of the most typical and influential main cultures of Chinese civilisation, the Yangtze River culture has played a leading role in the rejuvenation of national culture. Knowledge Graphs (KGs) and other technologies for semantic retrieval can be employed to enhance the efficiency of querying Yangtze River cultural data, thereby supporting the protection and inheritance of Yangtze River culture. Therefore, this paper designs a multimodal data-intelligent retrieval model based on knowledge graphs (KGs) and deep learning (DL). Build a knowledge graph (KG) for multi-modal data to gain a deeper understanding of the data and perform efficient information retrieval and knowledge discovery. In response to the problems of complex Chinese character shapes and semantic information in the narrative of the Yangtze River culture, this paper takes the Chinese pre-trained language model ChineseBERT as the semantic embedding layer of the text, integrates glyph and pinyin information, and improves the performance of traditional semantic parsing methods in the subtasks of entity mention recognition and relationship prediction; at the same time, a Convolutional Neural Network (CNN) is used to extract features from image data. According to the above simulation experiments, the model proposed in this paper has good intelligent retrieval accuracy.*

KEYWORDS: *Narrative of Yangtze River culture; Knowledge graph; Intelligent retrieval; Deep learning*

1 Introduction

With the rapid development of information technology, a large amount of unstructured data has been created, and as a result, problems in cultural management have also arisen [1]. Yangtze River culture is a typical representative symbol of the Chinese nation and an icon of Chinese civilisation; at the same time, it serves as a fundamental source of educational material for the socialist core values curriculum [2]. Knowledge graphs (KGs) as a type of semantic network are now one of the main reasons for the development of artificial intelligence (AI) in the era of big data [3]. KG is a relatively large-scale semantic network [4]. It has the components of entities, attributes and relationships, is based on large-scale knowledge, uses natural language processing (NLP) and other technologies to analyse and extract input knowledge, builds graph structures to show the relationships among this knowledge, and achieves more accurate and diverse semantic understanding and search for different industries and application scenarios [5]. If KG is viewed as a form of organised data for the real world, it can be queried, analysed statistically and mined [6]. KG is considered to be knowledge; semantic understanding, interpretation and reasoning can thus be carried out [7].

*200608023@jcut.edu.cn

https://doi.org/10.65102/is2026945 Published: 30 April, 2026

Convert the data of the Yangtze River Cultural Archives into KG for storage and display to reduce the workload of archive managers and improve the efficiency of user archive retrieval [8]. KG can represent the culture of the Yangtze River as a knowledge object and employ techniques such as NLP and knowledge extraction to extract specific entities, concepts and other information from this culture [9]. KG can annotate the Yangtze River culture at a fine-grained level, add more semantic information to it, and then use this semantic information in the downstream tasks of retrieval and recommendation. Thus, the model will be more effective in practice and convenient for users [10]. However, with a large amount of data available now, we need to explore the knowledge hidden in archived data and build a rich corpus of "Yangtze River culture" for retrieval, as well as advance knowledge-search technology [11]. When the user enters the search query for data retrieval, there is a mismatch with the actual data and therefore fails to retrieve it [12]. At the same time, with the complexity of the data, users cannot quickly obtain the relevant information about the data source [13].

Most of the existing research on KG construction is based on text data, and little attention has been paid to multimodal data such as images, videos and audio; therefore, research on the construction of multimodal KGs is relatively scarce [14]. With the development of deep learning (DL) and natural language processing (NLP), the accuracy of question-and-answer systems in solving complex problems has increased significantly, diverse expressions of knowledge have been used, and intelligent retrieval supporting multiple rounds of conversation has been achieved [15]. With the development of Large Language Models (LLMs), they are now capable of understanding and responding to complex semantic problems accurately, thus supporting the construction of intelligent question-and-answer and text-generation systems [16]. This paper selects the narrative scene of the Yangtze River culture as its subject and designs a multimodal intelligent retrieval model based on knowledge graphs (KGs) and deep learning (DL). Construct multimodal knowledge graphs to deepen the semantic understanding of data and improve the efficiency of information retrieval and knowledge discovery; Use ChineseBERT for text to enhance entity mention recognition and relationship prediction by combining glyph and pinyin features; Introduce Convolutional Neural Networks (CNN) to extract visual features from images and achieve collaborative modeling of graphics and text. The innovations are as follows:

(1)This article is the first to systematically construct a domain KG that integrates heterogeneous data from multiple sources such as text and images, enhancing the ability to express cultural semantics in a structured manner and perform associative reasoning.

(2)This article embeds glyph and Pinyin features based on ChineseBERT, effectively alleviating the problems of complex glyph and ambiguity in ancient literature, and significantly improving the performance of entity recognition and relationship extraction.

(3)This article integrates the semantic constraints of KG into the DL retrieval process, achieving alignment and interaction of text and image features in a unified semantic space, and improving the precision and interpretability of cross modal retrieval.

First, this paper will introduce the background of the study, and then list some applications of knowledge graphs (KGs) and deep learning (DL) in intelligent retrieval. The core of this paper is a multimodal data-intelligent retrieval model and its full application to the Yangtze River Cultural Data Mining problem. Several comparisons have been conducted in this paper to verify that the model performs well and have shown to be reliable. Finally, the conclusion part of this study will present a brief summary of the main results, identify the current deficiencies in the work, and suggest future research directions.

2 The Application of KG and DL in Intelligent Retrieval

2.1 KG

With the development of technology, the main contents of KG technology have gradually covered the flexible integration of heterogeneous data sources, description of data correlations, implementation of entity links, and large-scale knowledge reasoning. The general components of a KG construction are typically named entity recognition, relation extraction, entity alignment, knowledge reasoning, etc. Build a reasonable knowledge structure to help organise the data from various sources more easily, and perform deep semantic association and information fusion among them. To continuously add, improve and extend the Knowledge Graph (KG), and to address many deficiencies and inaccuracies in the large-scale open KG, KG completion technology can be employed. KG completion can continuously add new correct knowledge to incomplete KGs, and can also infer missing parts of triplets based on existing KGs; this is also known as link prediction of head and tail entities or relationships. The first is entity extraction in knowledge extraction, and it is a typical basic problem in natural language processing. Entity extraction is to extract segments with specific meanings from text to find entities in that domain in the text, extract atomic information from the text, and form entity nodes.

2.2 DL

Cross-modal retrieval is a typical use case of cross-modal learning that trains on information from several modalities but performs inference with only one modality. Use the first type of technology to extract information from one mode and convert it into another mode for cross-modal communication. DL deep learning refers to the learning of sample features by deep neural networks for automatic analysis and recognition of text, images and speech. Knowledge Graph construction based on deep learning can simultaneously parse multiple types of data, such as documents, videos and images, and extract information from the parsed data as knowledge for knowledge fusion and unification.

BERT can learn better-refined and more detailed semantic representations of textual data through the attention mechanism in Transformers and build a bidirectional model of context. Unsupervised learning can be used to pre-train BERT models on large-scale archival corpora for general semantic learning in the representation of archival texts; then, in conjunction with specific classification tasks, the model parameters are fine-tuned using a small amount of annotated data to obtain domain-specific semantic representations to support downstream archive classification tasks. Traditional manual feature extraction methods are unable to describe the semantic content of image data effectively, but CNNs can be used for image classification and have achieved good results. The structure of the CNN is as follows: Figure 1.

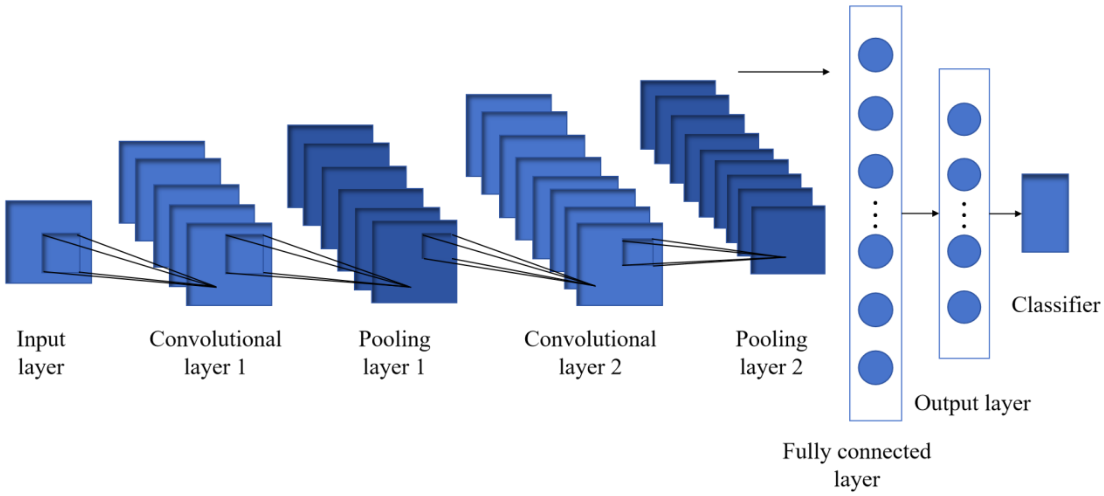

Figure 1: CNN Structure

3 A Multimodal Data Intelligent Retrieval Model Combining KG and DL

3.1 Model Construction

Chinese, as a symbol system, has many different meanings and expressions, and its characters can also be used to represent various ideas. Therefore, glyph information can be used to strengthen the effect of Chinese knowledge Q&A. Pinyin for Chinese can be used to express semantic and grammatical information, and the same characters have different meanings when spoken in different ways. Therefore, ChineseBERT is employed as the semantic embedding layer of text in this paper and integrates glyph and pinyin information. Based on the above, this paper introduces a relationship prediction method based on ChineseBERT TextCNN Softmax. The input part of the ChineseBERT model adds a learnable fusion vector that includes character vectors, glyph vectors, and the corresponding character pinyin vectors. Integrate glyph and Pinyin information in the ChineseBERT model to better capture the semantic features of Chinese characters.

However, the previous methods for classification fail to adhere to the label-ordering constraints of the entity recognition task during the prediction phase. Therefore, this paper added a CRF layer downstream of ChineseBERT. The CRF layer learns the dependency of the labels and modifies the predicted label sequence globally to obtain a label sequence that satisfies the rules of entity recognition. The goal of relation prediction is to semantically match the relationship in the problem text with that in a knowledge graph (KG) and select the most suitable relationship. This paper models the relationship prediction problem as a semantic matching problem: given the problem text and candidate relationship descriptions, both are encoded by semantic encoders, and then their matching probabilities are calculated; among all candidate relationships, the one with the highest matching probability is selected as the final prediction result. This way does not need to adhere to the limitations of traditional classification methods, and it can also handle models for open-domain tasks or those with a long tail. The structure of the model is as follows: Figure 2.

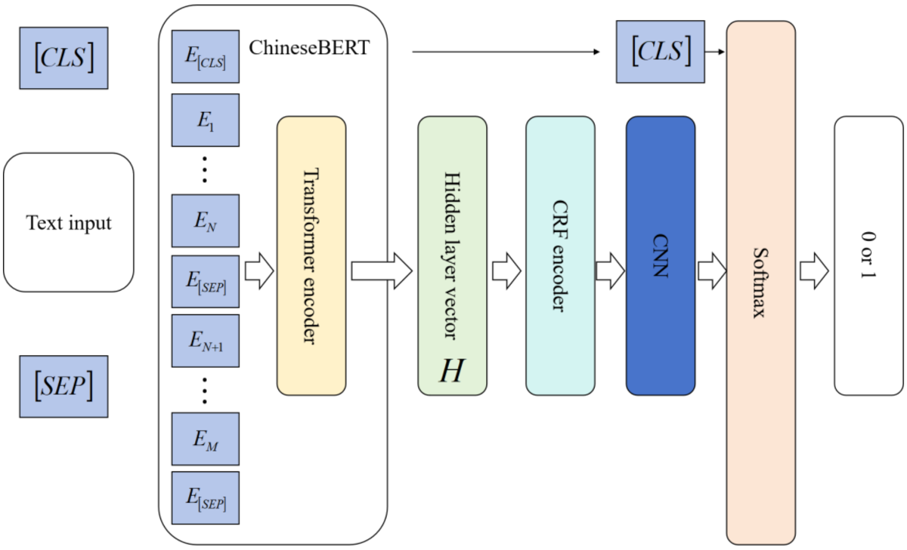

Figure 2: Model Architecture

3.2 Algorithm Principle

This article adopts a sequence annotation architecture combining ChineseBERT and CRF (ChineseBERT-CRF) for joint entity and relationship extraction. Embed the problem text $x=(v_1,v_2,\cdots,v_n)$ into the ChineseBERT pre trained language model to obtain its context sensitive hidden layer representation sequence H, which is used to represent the semantic and contextual information captured by the text. Next, these hidden layer vectors H are input into the CRF layer, which processes the input feature sequence and calculates the annotation path with the highest conditional probability, thus obtaining the predicted label sequence. The method expression for calculating the score of the input sequence y and the predicted label sequence DD in the CRF layer is:

$$H = ChineseBERT(x) \tag{1}$$

$$s(x,y)=\sum_{i=0}^{n} A_{y_i,y_{i+1}}+\sum_{i=1}^{n} H_{i,y_i} \tag{2}$$

$$L=\ln\left(\sum_{y'} e^{s(x,y')}\right)-s(x,y) \tag{3}$$

In the formula: A is the transition matrix of the CRF layer; $A_{y_i,y_{i+1}}$ is the score obtained by transferring from prediction label y_i to prediction label y_{i+1}; H_{i,y_i} is the score obtained by marking the i word as the predicted label y_i. To calculate the score of this label sequence, the sum of $A_{y_i,y_{i+1}}$ from i to $0\sim n$ and H_{i,y_i} from i to $1\sim n$ are added together.

Extract feature vectors from the image data using a pre-trained deep CNN, and the feature extraction process is as follows:

$$V_{image} = CNN_{features}(I) \tag{4}$$

Among them, V_{image} is the feature vector of image I ; $CNN_{features}$ is the feature extraction layer of CNN.

When the input sentence contains multiple entities, the multiple triplets generated by entity linking may introduce semantically irrelevant or redundant knowledge. If directly integrated into sentence modeling, it is easy to introduce noise in subsequent self attention calculations, which can interfere with the original semantic expression. To this end, this article designs a visual layer mechanism that explicitly constrains the attention range by constructing a visual matrix M : only allowing information exchange within triplets and prohibiting cross regional attention propagation between triplets and other parts of the sentence, effectively isolating external knowledge interference and ensuring the semantic integrity of the original sentence. This mechanism enhances the controllability of knowledge injection while improving the model's ability to focus on core semantics. The definition of visualization matrix M is as follows:

$$M_{ij} = \begin{cases} 0 & w_i, w_j\, same\ branch \\ -\infty & else \end{cases} \tag{5}$$

In the formula, w_i is the i word in the input sentence, and w_j is the j word in the input sentence.

In the relationship prediction task, first add special identifiers $[CLS]$ (beginning of sentence) and $[SEP]$ (end of sentence) to two input texts respectively, forming a standard input sequence; Then input these two texts into the ChineseBERT CRF network for embedding semantic information. ChineseBERT CRF encodes the text to obtain the hidden layer vector H and $[CLS]$ vector y_1, and then passes the hidden layer vector H into TextCNN, which includes three convolutional layers with step sizes of 1, 3, and 5, respectively. Through these convolutional layers, features can be extracted to obtain vector y_2. Finally, vector y_2 and y_1 are concatenated and input into the Softmax layer for classification. The Softmax layer is a commonly used multi class regression model used to predict the probability values of binary labels, and the output binary label values may be 0 or 1. For a sample input feature x with K category label values, the prediction probability calculation formula is:

$$p(y = k \mid x) = \frac{\exp(W_k x)}{\sum_{k=1}^{k} \exp(W_k x)} \tag{6}$$

In the final answer generation stage, this article collaboratively constructs answers based on entity mention recognition and relationship prediction results. To achieve precise entity linking, concatenate the description of candidate entity e_i in question text x and KG into a matching pair, input it into ChineseBERT Softmax semantic matching model, calculate its matching score and normalize it to link probability:

$$P_i^e = ChineseBERT\_Soft\max([x; e_i]) \tag{7}$$

In the prediction process, for each candidate entity e_i and candidate relationship r_{ij}, the formula for calculating the score is:

$$Score = \gamma P_i^{e(T)} + (1-\gamma) P_{ij}^{r(T)} \tag{8}$$

Among them, γ is an adjustable hyperparameter used to dynamically balance the contribution of entity matching and relationship matching.

4 Result Analysis and Discussion

The experimental data in this paper are authoritative cultural institutions across the Yangtze River Basin, such as digital collections from museums and archives, digital records of intangible cultural heritage, and geographic and cultural images along the Yangtze River; these cover two modes: high-definition images and multi-source text. The experimental platform is a high-performance workstation that has been set up; it includes an Intel Core i7-14700K processor (20 cores, 28 threads) and an NVIDIA GeForce RTX 3080 GPU (10GB video memory) to ensure high-efficiency execution of model training and inference; A Python 3.9-based PyTorch 1.10.0 deep learning framework is used for the software environment, and CUDA 11.3 accelerated computing is employed to provide stable and reliable operational support for multimodal KG construction and semantic retrieval models.

To assess the performance of our model in the multimodal intelligent retrieval task for Yangtze River culture, a comparative experiment was set up, and BiLSTM CRF and K-BERT were chosen as the baseline methods for fair comparison on a self-built test set of Yangtze River culture digital art resources. Retrieval precision was selected as the index of comparison in the experiment, and as shown in Figure 3, the model presented here performed better than the baseline model in all the tests; thus, it was confirmed to be superior. This paper integrates the structured prior of KGs with the semantic generalization capability of DL in the model; uses ChineseBERT CRF for end-to-end joint recognition of entity mentions and relations in text to mitigate error propagation effectively; innovatively introduces a ChineseBERT TextCNN Softmax hybrid structure to enhance the robustness of relation discrimination significantly; at the same time, CNN branches perform multi-level feature extraction on images to support cross-modal semantic alignment. In short, this model can improve the accuracy of cultural semantic understanding and the efficiency of retrieval simultaneously.

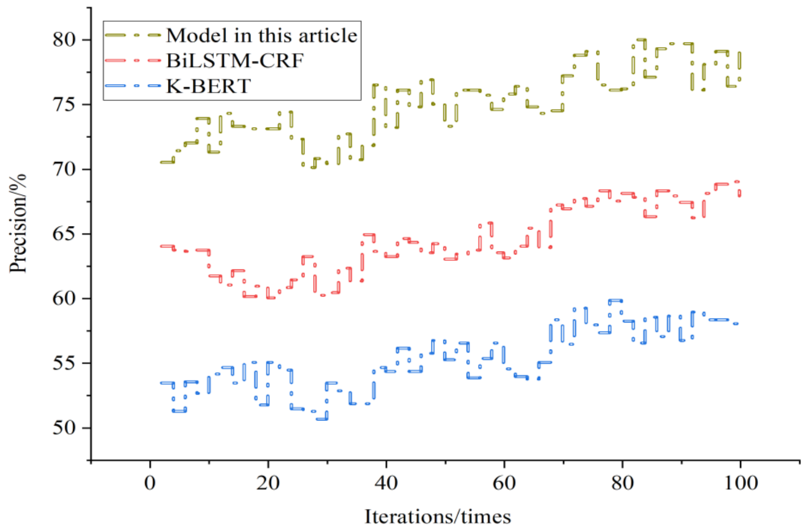

Figure 3: Precision Comparison

To test how fast our model runs in the Yangtze River cultural intelligent retrieval task in real time, BiLSTM CRF and K-BERT were also evaluated under the same hardware conditions and using the same test set. The results are as follows: Figure 4 Based on the experimental results shown in this paper, the entire curve of the model is below that of the baseline model, and thus it can be used for interactive retrieval in cultural applications with a low-latency requirement. This paper restricts the range of attention for triplets in the visual layer to prevent computational inflation due to KG embeddings; a ChineseBERT CRF module shares encoding layer parameters to jointly extract text entities and relations, thereby reducing pipeline cascading delay; and a TextCNN uses multi-scale one-dimensional convolution for parallel processing of hidden layer sequences to significantly reduce inference overhead. This model will be able to provide a reasonably responsive cultural intelligence service system at a high precision.

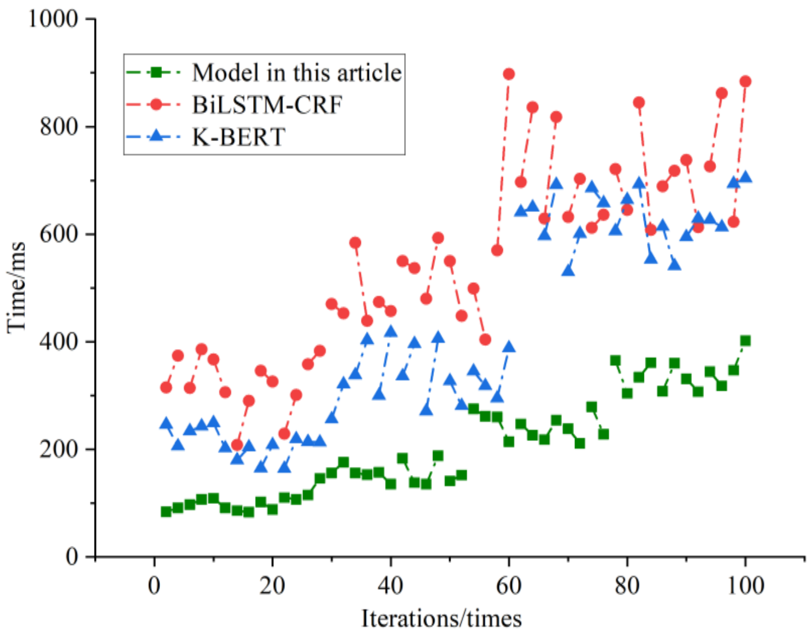

Figure 4: Time comparison

To comprehensively assess the extent to which the results of our model for intelligent retrieval of Yangtze River culture cover the entire content, a comparison experiment was conducted with BiLSTM-CRF and K-BERT using recall rate as the main indicator. The results are as follows: Figure 5. Based on the above results, it can be seen that the entire curve of the model in this paper is higher, and thus it will be more comprehensive and effective in covering relevant cultural resources to address the issue of missed detections. The multimodal KG built in this paper expands query intent and merges synonymous entities to expand the coverage of the search candidate set; ChineseBERT CRF conducts end-to-end entity relationship joint recognition to avoid the omission of relationships caused by entity recognition errors in traditional pipelines; ChineseBERT TextCNN Softmax improves the ability to distinguish complex and ambiguous relationships; and the image CNN branch supplements visual semantic clues to help recall cross-modal associations. In short, this model can help discover more cultural knowledge in an organized way at the same time.

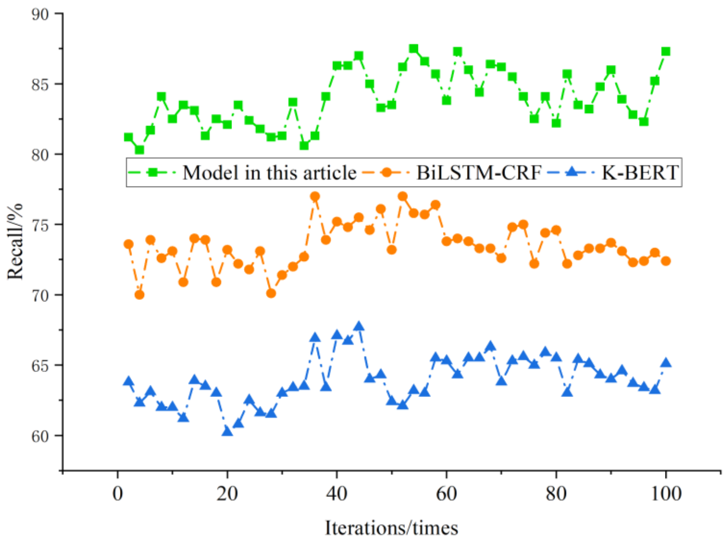

Figure 5: Comparison of Recall Rates

To assess the general performance balance of the model in Yangtze River cultural intelligent retrieval comprehensively, F1 score was chosen as the main indicator and compared with BiLSTM-CRF and K-BERT in the experiment. The results are as follows: Figure 6. Based on a self-built multimodal test set of Yangtze River culture, the three sub-tasks of this experiment are entity linking, relationship classification, and cross-modal retrieval. Based on the above experiments, the F1 scores of the proposed model in this paper surpass those of the previous models, showing good retrieval efficiency. The KG and DL deep fusion architecture in this model adds domain knowledge constraints to improve precision and uses data-driven generalisation ability to ensure recall coverage; the ChineseBERT CRF joint extraction module reduces error accumulation through end-to-end training and significantly enhances the joint F1 score of entity and relationship recognition. Semantic representation and local combination feature for long-tail and ambiguous relationship prediction in ChineseBERT TextCNN Softmax model improve discriminability globally.

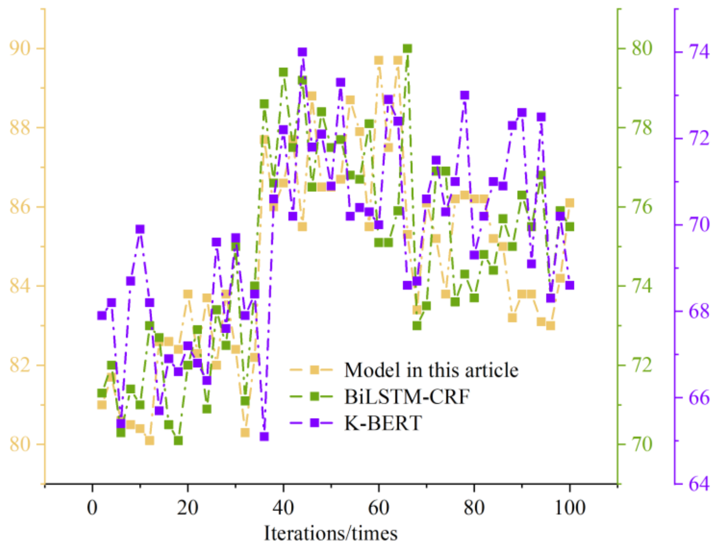

Figure 6: Comparison of F1 values

To assess the efficiency of computational resource utilization for our model in intelligent retrieval of Yangtze River culture, it was compared with BiLSTM CRF and K-BERT on a unified hardware platform. Monitor GPU memory usage, peak CPU memory, and unit query energy consumption, and display the results in Figure 7. According to the experiment results, the model presented in this paper has high precision and its entire curve is above that of the comparison model; therefore, it is confirmed that it is more resource-efficient and suitable for deployment in an edge-computing environment or a high-concurrency cultural service platform.

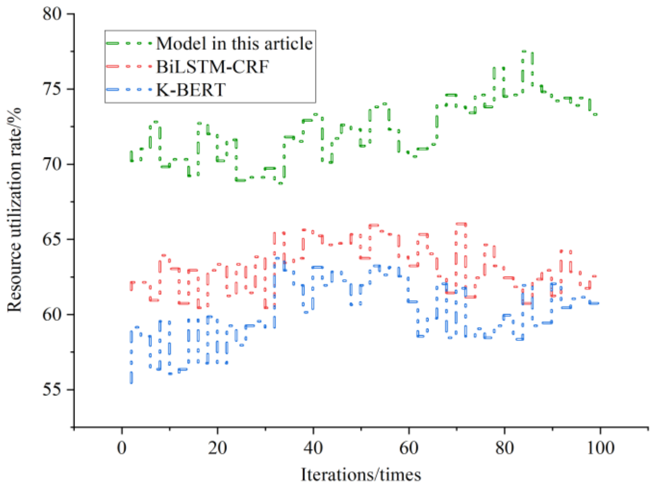

Figure 7: Comparison of Resource Utilisation Rates

To test the actual application effect of the model proposed in this paper for intelligent retrieval of Yangtze River culture, a user satisfaction comparison experiment was conducted. Fifty domain users were asked to complete fifty tasks using the model, BiLSTM-CRF, and K-BERT systems in a real retrieval scenario. Four dimensions were scored by the Likert 5-point scale: relevance, response speed, interface friendliness and accuracy of cultural expression. The results are as follows: Figure 8 Based on the above experiments, the user satisfaction of our model is relatively high. The model retrieval precision in this paper is higher, so the interference of irrelevant or erroneous results can be reduced; at the same time, the model in this paper has a shorter time consumption, a faster response speed, and can meet user needs.

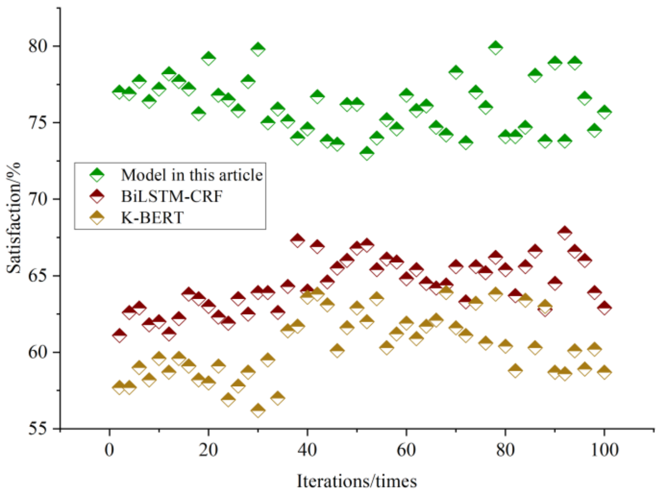

Figure 8: Satisfaction Comparison

5 Conclusion

A multimodal intelligent retrieval model combining KG and DL is proposed in this paper to improve the organization and accuracy of discovery for digital resources of Yangtze River culture. By building a domain multimodal KG that includes both text and images, the model can obtain a structured representation and deep association reasoning for cultural semantics; to address the problems of complex character forms and semantic ambiguity in Chinese ancient books and local literature, we have innovatively employed ChineseBERT with a triple enhancement of character-form pinyin semantics as the text encoder to significantly improve the robustness of entity mention recognition and relationship prediction, and simultaneously introduced CNN to extract visual features from images for alignment and complementarity of graphic and textual semantics in a unified space. A large number of simulation experiments have shown that this model is significantly better than the Bilstm-CRF and K-BERT baselines in terms of precision, real-time capability, recall integrity, user satisfaction, and other essential indicators; thus, its progressiveness technology and application effectiveness have been validated.

The model shown in this paper is also not perfect. Currently, KG primarily integrates text and images; rich-media forms such as videos, audio and 3D models have not been incorporated, and thus the full modelling of dynamic cultural processes remains to be achieved. In the future, more multimodal knowledge graphs will be introduced, cross-modal alignment pre-training and incremental update mechanisms will be added, and the construction of an intelligent service system for the digital resources of the Yangtze River Culture will be further promoted.

Acknowledgements

This work was supported by the 2025 Provincial Philosophy and Social Science Research General Project: Research on the Symbolic Narration and Innovative Design of Multi-ethnic Dance Costumes in the Wuling Mountain Area (No. 25Y133).

About the Author

Lili Zhang was born in Chengdu, Sichuan Province, China, in 1984. Doctor of Philosophy, Lecturer. She is a graduate of De La Salle University-Dasmarinas. She works at the College of Arts, Jingchu University of Technology, now. She is interested in the visual expression and dynamic inheritance of intangible cultural heritage, as well as aesthetic education, and so on.

Yu Fang was born in Jingmen City, Hubei Province, China, in 1984. She got her Ph.D. at De La Salle-Dasmarinas. She teaches in the College of Arts at Jingchu University of Technology now, and her specialization is Visual Communication Design. Her research on traditional culture and digital illustration.

Reference

[1] Bellini V, Di Sciascio E, Donini F M. A qualitative analysis of knowledge graphs in recommendation scenarios through semantics-aware autoencoders[J]. Journal of Intelligent Information Systems, 2024, 62(3): 787-807.

[2] Byeon H, Chunduri V, Narang G. Deep learning model for recommendation system using web of things based knowledge graph mining[J]. Service Oriented Computing and

Applications, 2025, 19(1): 57-76.

[3] Li C, Cao Y, Zhu Y. Ripple knowledge graph convolutional networks for recommendation systems[J]. Machine Intelligence Research, 2024, 21(3): 481-494.

[4] Lee S J, Ahn S, Lee E. Extended Knowledge Graph using Relation Modeling between Heterogeneous Data for Personalized Recommender Systems[J]. Smart Media Journal, 2023, 12(4): 27-40.

[5] Khan A, Ke X, Wu Y. Graph data management and graph machine learning: Synergies and opportunities[J]. ACM SIGMOD Record, 2025, 54(2): 28-42.

[6] Steenwinckel B, Vandewiele G, Weyns M. INK: knowledge graph embeddings for node classification[J]. Data Mining and Knowledge Discovery, 2022, 36(2): 620-667.

[7] Liu W. Study on complement of knowledge map of educational resources based on semi-supervised learning[J]. International Journal of Business Intelligence and Data Mining, 2025, 26(1-2): 100-118.

[8] Renzi G, Rinaldi A M, Russo C. A storytelling framework based on multimedia knowledge graph using linked open data and deep neural networks[J]. Multimedia Tools and Applications, 2023, 82(20): 31625-31639.

[9] Li S, Wang J, Xu Z. Automated compliance checking for BIM models based on Chinese-NLP and knowledge graph: an integrative conceptual framework[J]. Engineering, Construction and Architectural Management, 2025, 32(6): 3832-3856.

[10] Tao S, Qiu R, Cao Y. Path-guided intelligent switching over knowledge graphs with deep reinforcement learning for recommendation[J]. Complex & Intelligent Systems, 2023, 9(6): 7305-7319.

[11] Tran T K, Ta C D C, Phan T T. Simplified effective method for identifying semantic relations from a knowledge graph[J]. Journal of Intelligent & Fuzzy Systems, 2022, 43(2): 1871-1876.

[12] Chen Y, Wu L, Zaki M J. Toward subgraph-guided knowledge graph question generation with graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 35(9): 12706-12717.

[13] Ahmad R A, Rehm G. Knowledge injection for field of research classification and scholarly information processing[J]. Computación y Sistemas, 2024, 28(4): 2153-2169.

[14] Obono B M, Liu W. Knowledge Graph Question-Answering Method Based on Neural Networks[J]. Open Access Library Journal, 2025, 12(4): 1-20.

[15] Wu D, Tang M, Zhang S. KPRLN: deep knowledge preference-aware reinforcement learning network for recommendation[J]. Complex & Intelligent Systems, 2023, 9(6): 6645-6659.

[16] Huang Z, Lin Z, Gong Z. A two-phase knowledge distillation model for graph convolutional network-based recommendation[J]. International Journal of Intelligent

Systems, 2022, 37(9): 5902-5923.