ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Research on the Construction Method of Guanzhong Folk Habitat Cultural Gene Knowledge Graph by Integrating Multimodal Data

Hao Zhang[1,*]

[1] Department of Architectural and Environmental Art, Xi'an Academy of Fine Arts, Xi'an, Shaanxi, 710065, China

SUMMARY: *This paper explores the method of constructing a genetic knowledge map of Guanzhong residential culture by integrating multimodal data in order to reorganize the knowledge of Guanzhong residential culture. The study firstly proposes four entity types in terms of the content structure of Guanzhong residents' culture: deity culture, Confucianism culture, dwelling culture and aesthetic concept. Then the identification and extraction process of Guanzhong residential culture genes was designed, and the visual features of Guanzhong residents' culture were extracted based on the selected text data and picture data. Finally, Bert-wwm-ext-BiLSTM-Attention-CRF and BiLSTM-PCNN-Attention are selected as the entity recognition model and the relationship extraction model, respectively, to construct the Guanzhong residential culture gene knowledge graph. The knowledge graph construction method selected in this paper is validated, and the results of multi-model comparison experiments show that the entity recognition accuracy of Bert-wwm-ext-BiLSTM-Attention-CRF reaches more than 70%, and the relationship extraction performance of BiLSTM-PCNN-Attention has a better performance advantage among all the compared models, and this paper Both models designed are suitable for building Guanzhong residential culture gene knowledge map.*

KEYWORDS: *Multimodal data; Cultural genes; Knowledge graph; Bert-wwm-ext-BiLSTM-Attention-CRF; BiLSTM-PCNN-Attention; Guanzhong folk house culture*

1 Introduction

The “Guanzhong” region of Shaanxi is known as the “800-mile Qin River”. It has been proven that the earliest houses in China were built on this land. After thousands of years of changes, Guanzhong houses have become a school of their own in China's residential architecture with their own unique simple and grand architectural style [1]. The Guanzhong houses are generally characterized by compact layout, economical land use, strict selection of materials and construction quality, flexible treatment of indoor and outdoor space, and high level of decorative art, which are valuable heritage of Chinese architectural culture [2-4]. However, with the process of modern urbanization, modern Guanzhong residential architecture has been influenced by foreign building materials and building types, and the residential building types are becoming more and more popularized [5, 6].

Knowledge mapping, as a powerful tool for knowledge representation and organization, provides new ideas and methods for the study and inheritance of Guanzhong folk dwelling culture [7, 8]. By organizing the information and cultural relics of folk dwelling culture,

*zhanghao@xafa.edu.cn
https://doi.org/10.65102/is2026264 Published: 30 April, 2026

combining the classification and evolutionary relationship of Guanzhong dwelling culture in different regions and periods, and constructing the genetic knowledge map of Guanzhong dwelling culture, the scattered knowledge of dwelling culture can be integrated to form a structured knowledge network [9-11]. And in this process, by fusing multimodal data such as text, image, audio and video related to Guanzhong Minjuwen, it can describe the entities and their relationships more comprehensively, which is of great value for with the promotion of the inheritance and protection of Guanzhong Minjuwen culture, the promotion of Guanzhong Minjuwen culture research and innovation, and the enhancement of the dissemination effect of Guanzhong Minjuwen culture [12, 13]. In the context of the rapid development and application of artificial intelligence and knowledge graph, the construction of Guanzhong folk house culture gene knowledge map is of great significance for the protection and inheritance of Guanzhong folk house culture.

This paper provides new research ideas and technical methods for the construction of cultural gene maps. The selected entity recognition model (Bert-wwm-ext-BiLSTM-Attention-CRF) replaces the Bert model with the Bert-wwm-ext model on the basis of the deep learning Bert-BiLSTM-CRF model, and at the same time introduces the attention mechanism, which is applied to the Guanzhong residents' cultural entity recognition to enhance the model's ability to recognize graphical and textual features in the cultural domain of Guanzhong residents through the attention mechanism enhances the model's ability to recognize the graphic and textual features of Guanzhong residents' cultural domain, and effectively extracts the information of Guanzhong residents' cultural entities. The selected relationship extraction model (BiLSTM-PCNN-Attention) introduces BiLSTM on the basis of the deep learning PCNN-Attention model, and applies it to Guanzhong residential culture relationship extraction, optimizes the understanding of graphic sequences and the ability to capture contextual information through BiLSTM, and improves the model's accuracy for the extraction of the relationship between entities. The accuracy of the model is improved by optimizing the understanding of graphical sequences and the ability to capture contextual information.

2 Construction of Guanzhong Folk Habitat Cultural Gene Knowledge Graph by Integrating Multimodal Data

2.1 Research on Guanzhong Folk Houses Cultural Entities

From the initial economic activities of society to the establishment and development of social relations and superstructure, folk culture can be seen in miniature throughout the process, and folk culture is complex. As the carrier of folk culture, folk houses, like culture, have always been present in the development of the society. The folk culture of Guanzhong region can be divided into the following categories:

(1) Spiritual Culture Embodied in Folk Houses

In the traditional Guanzhong folk houses, there are some common deities, such as the Land God, who blesses “living in peace and giving to the people and making them happy”. The God of Zao Wang, who brings good fortune. The God of Door, who drives away evils and gives peace to the whole family. In addition, the more popular folk god of wealth, the god of grain, the three stars of happiness, fortune and longevity, the goddess of mercy and so on.

(2) Confucian Culture Embodied in Folk Houses

The common feature of the dwellings in Guanzhong region is the integration of Confucianism, Taoism and patriarchal culture into the architectural field. Since ancient times, the people in Guanzhong have been influenced by the Confucian and Taoist ideas of "valuing the root and cherishing life", and regard "offering sacrifices to heaven, earth and ancestors" as the basic principles of the ritual system. In architectural styles and designs, there are ritual buildings for sacrificing to heaven, earth and ancestors. In the dwellings, most of them retain the sacrificial spaces of "respecting heaven and offering sacrifices to ancestors". During the Qingming Festival and other festivals, people kneel down to pay homage to their ancestors and express deep sorrow and reverence.

(3) Living room culture

The words on the traditional buildings in Guanzhong not only send people's good expectations for the future, but also have a very obvious decorative meaning. On the door pillars, window frames, plaques and couplets of the houses, words such as “Wan”, “Quan” and “Fu” are usually used, so that the buildings have strong humanistic characteristics in addition to artistic characteristics. This makes the buildings not only have artistic characteristics, but also have strong humanistic characteristics, which realizes the perfect combination of architecture and humanities, and makes the houses truly realize the harmony and unity of people and architecture.

(4) Aesthetic concept

The artistic characteristics embodied in Guanzhong's residential buildings have strong aesthetic principles. Compared with other buildings, Guanzhong residential buildings are close to the golden section in the ratio of face width and depth on the plan, which fully shows the proportion and structure of Guanzhong residential buildings, and the tiles on the eaves, the rafters under the eaves, and the arches on the beams are all unique and orderly, which constitutes the unique modeling art and decorative rhythms of residential buildings, and forms the unique charm of Guanzhong residential buildings.

2.2 Identification and Extraction of Cultural Genes of Guanzhong Folk Houses

The process of recognizing and extracting the cultural genes of Guanzhong folk houses is shown in Figure 1. The identification and extraction of cultural genes are divided into three steps. The first step is to collect Guanzhong residents' culture-related data through literature review, field research and network search, screen out scientific, complete and high-quality Guanzhong residents' culture data, and convert the format to get text data and image data. In the second step, artificial intelligence technology is used to determine the labels for the identification and extraction of Guanzhong residents' culture, and to clarify the categories of cultural genes to be extracted. In the third step, the keywords are extracted from the text data by using the large language model, the features of the picture data are recognized and classified by using convolutional neural network, and finally the visual features of Guanzhong residents' culture are extracted.

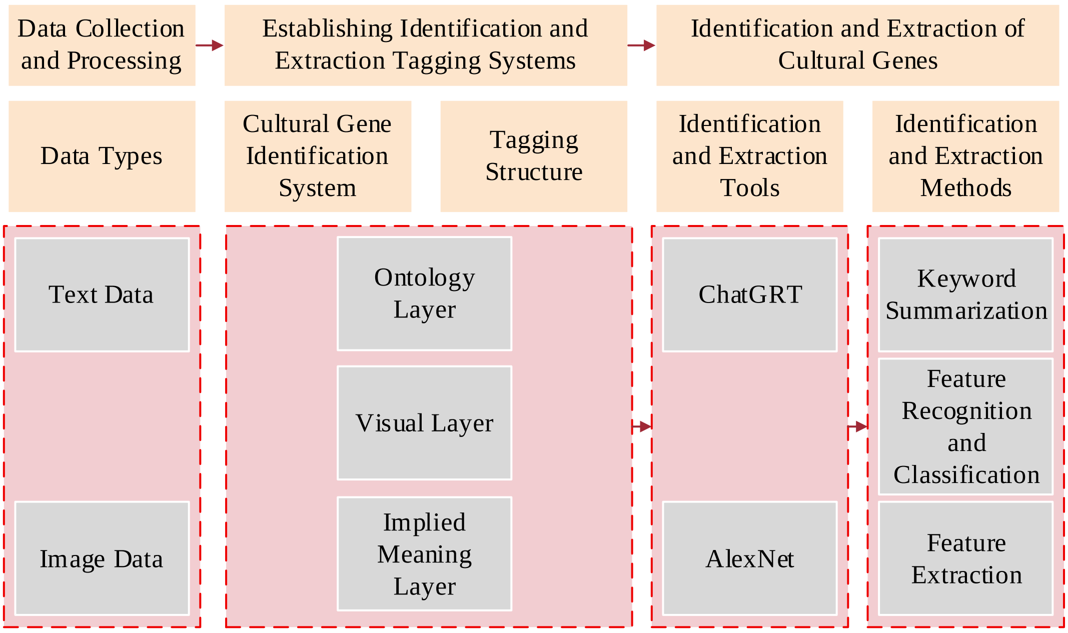

Figure 1: Identification and extraction process of Guanzhong residential cultural genes

(1) Data sources

The specific textual data based on which the cultural genes of Guanzhong folk houses are extracted are mainly referred to ancient literature and local records, such as the records of geography and customs of Guanzhong in the early historical books such as the Shiji, the Han Shu, and the Shui Jing Ji. Specific picture data come from Guanzhong folklore atlases published by local cultural departments and publishing houses.

(2) Labeling of cultural genes of Guanzhong residents

The identification of cultural genes is a prerequisite for the extraction of cultural genes of Guanzhong residents, and the extraction of cultural genes of Guanzhong residents in this paper mainly relies on the ability of the large language model to deal with the summarization task. Therefore, in order to ensure the scientific nature of the extracted content, it is necessary to establish a cultural gene labeling system. The cultural gene labels of Guanzhong residents extracted in this paper are shown in Table 1. Among them, the visual layer is mainly extracted to the color system, grain pattern, material texture, spatial proportion, stylistic features and decorative techniques about Guanzhong residents' culture.

Table 1: Guanzhong cultural gene extraction label

Structural hierarchy	Extraction label
Body layer	Cultural entity and material carrier
Visual layer	Color system
	Pattern pattern
	Material texture
	Spatial ratio
	Molding feature
	Decorative technique
Hidden layer	Cultural connotation

(3) Extraction of cultural genes

1) Extraction of text data

Large language model is a natural language processing technology based on deep learning, which can understand and generate human language, analyze a large amount of text data, and learn the structure, semantics, and contextual relationships of language. The use of large language models for cultural gene recognition and extraction firstly preprocesses the data and transforms it into a format that can be recognized and used by computers, and secondly allows the accuracy of the results of the large language model summarization task to be guaranteed by means of given extracted keywords to achieve the extraction of cultural genes from the text content. In this paper, we use ChatGPT summary task ability to extract textual information. After inputting the keywords to be extracted and the corresponding cultural gene types in the ChatGPT dialog box, ChatGPT will automatically identify the genes such as the color system, grain pattern, material texture, spatial proportion, styling features and decorative techniques in the textual paragraphs.

2) Extraction of image data

The artificial neurons in the convolutional neural network layer are sensitive to edge information and have the ability to migrate features by simulating neurons in the visual area, which can extract the features in the image and thus improve the classification accuracy of the image data. The use of convolutional neural network for the identification and extraction of cultural genes, the first step is to train the classification model, using the convolutional neural network model to learn the features of the image, and then the image data can be predicted by using the trained model, which can complete the classification of the cultural genes of the image. In this paper, a pre-trained AlexNet model is used to recognize images in a given folder and output the model's prediction of the category to which the image belongs and its corresponding probability. The graphical representation requires a symbolic presentation of the object of study, and the use of simple lines to show the artistic characteristics of the culture of Guanzhong residents is a prerequisite for graphical representation. In this part, the PIL library is used to read the images, and the binary image is generated through thresholding, and the “get_contour” function is called to recognize and extract the shape characteristics of Guanzhong residents' rooms from the binary image.

2.3 Guanzhong residential culture gene knowledge map construction

2.3.1 Knowledge map construction methods

Building a knowledge graph is usually divided into two parts: named entity recognition and relationship extraction. Among the mainstream methods are entity and relationship extraction methods based on RNN model, CNN model and LSTM model.

(1) Entity and relationship extraction methods based on RNN models

The RNN model is called recurrent neural network, which can make good use of the relationship between sequences, and can be good for input sequences with continuity.The structure of the RNN network is shown in Fig. 2, in which W, U, V denote the weight matrix, and f denotes the nonlinear activation function.

The formula of RNN is shown below:

$$h_t = f\left(Wh_{t-1} + Ux_t\right) \tag{1}$$

$$y_t = Soft\max\left(Vh_t\right) \tag{2}$$

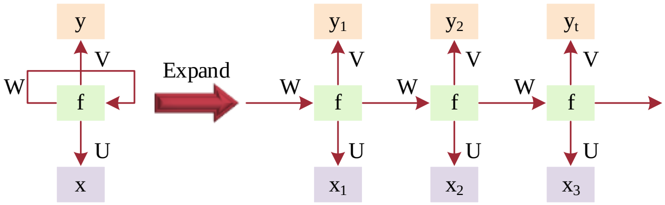

Figure 2: RNN structure network diagram

However, RNN is prone to the problems of gradient vanishing, gradient explosion and long training cycle in training. Therefore CNN model based entity and relationship extraction methods are proposed.

(2) CNN model based entity and relationship extraction method

CNN model is called convolutional neural network structure is relatively simple, in the network structure of CNN contains the front convolutional layer and the back full connectivity layer can be learned in parallel, compared with RNN training speed is faster.

(3) Entity and Relationship Extraction Method Based on LSTM Models

LSTM is a kind of RNN. The LSTM model is based on RNN with the addition of forgetting gates, input gates, unit states and output gates. During the training process, information is added or removed through the gate structure and the history of the context is selectively forgotten, updated and passed on.The LSTM model effectively prevents gradient explosion and gradient vanishing present in the RNN, while better capturing dependencies over longer distances.

The forgetting gate of LSTM is shown in equation (3):

$$f_t = \sigma\left(w_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{3}$$

The update gates of LSTM are shown in equations (4) to (6):

$$i_t = \sigma\left(w_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{4}$$

$$\tilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{5}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{6}$$

The output gates of the LSTM are shown in equations (7) to (8):

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{7}$$

$$h_t = o_t * \tanh\left(C_t\right) \tag{8}$$

This separated framework model makes extraction tasks easier to handle and each component can be more flexible. But there can be some common problems: the correlation between the two subtasks is ignored. Each sub-task is a separate model, and the results of entity recognition may affect the performance of relationship classification and lead to error transmission, as well as the problem of no correlation between extracted entities, leading to redundant information. Unlike the pipeline approach, the joint learning framework utilizes a single model to extract entities and relationships. It can effectively integrate the information of entities and relations to get the triad directly and achieve better results. According to the different modeling objects, joint extraction can be further divided into parameter sharing methods and sequence labeling methods.

Knowledge graph as a kind of semantic network, through the nodes in the graph to represent the entities, using the edges between the nodes to represent the various semantic relationships between the entities, and it plays a great role in the development of relationships, entity exploration and other applications. Benefiting from the rich semantic information of the knowledge graph, the model can be made to better understand the natural language through the knowledge graph, and further better understand the user's intention. Utilizing the strong feature extraction capability of deep learning to achieve knowledge extraction naming entity recognition, entity relationship extraction and other tasks in the knowledge graph construction process makes the knowledge graph achieve greater development with the help of deep learning technology. In this regard, this paper selects Bert-wwm-ext-BiLSTM-Attention-CRF as the entity recognition model and BiLSTM-PCNN-Attention as the relationship extraction model, respectively, to build Guanzhong folk house cultural gene knowledge graph.

2.3.2 Guanzhong Folk Houses Cultural Gene Knowledge Mapping Construction Methods

This study proposes the construction method of Guanzhong residents' cultural multimodal knowledge map for Guanzhong residents' cultural graphic data, and the research framework of the construction method is shown in Figure 3. First, the modal layer is designed from the content and modal dimensions of Guanzhong residents' cultural items, and the modal types of Guanzhong residents' cultural digital resources are expanded. Second, automatic knowledge extraction for Guanzhong residents' cultural graphic digital resources is realized. Finally, a multi-scene-driven graphic knowledge linking scheme is proposed to realize multi-modal interaction. Through the above three aspects of research, the construction of multimodal knowledge graph for Guanzhong residents' cultural graphic data is basically realized.

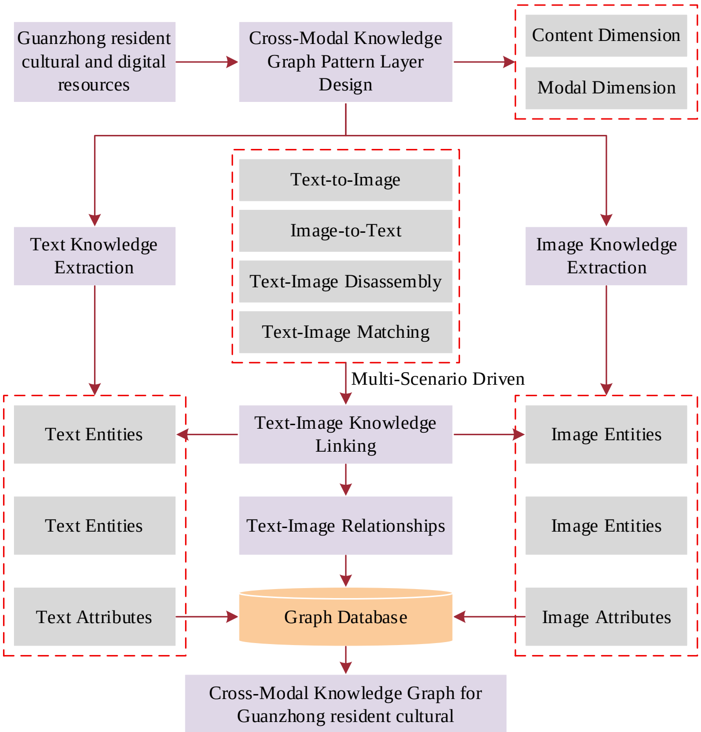

Figure 3: The construction method of the multi-modal knowledge map

(1) Design of multimodal knowledge mapping schema layer

This study standardizes the multimodal knowledge mapping schema layer of Guanzhong residents' culture from the content and modal dimensions of Guanzhong residents' cultural items, which can lay the foundation for the subsequent work of Guanzhong residents' cultural graphic and modal knowledge extraction. From the analysis of the content dimensions of Guanzhong residents' cultural items, this study, by referring to the conceptual model in the field of cultural heritage and existing studies, proposes four entity types of Guanzhong residents' cultural content dimensions, namely, deity culture, Confucian culture, house culture and aesthetic concept, based on which the different Guanzhong residents' cultural items can be expanded to include entity subtypes. Analyzing the modal dimension of Guanzhong residents' cultural projects, the text modal resources include the content entities of Guanzhong residents' cultural projects and the entities of keywords frequently appearing in Guanzhong residents' cultural texts, and the image modal resources are divided into the overall visual entities and the target visual entities based on the granularity of the images.

Guanzhong residents' cultural program content dimension inter-entity relationship mainly includes the relationship between entities of the four Guanzhong residents' cultural program content dimensions mentioned above, and the specific Guanzhong residents' cultural program needs to be expanded and deepened on the basis of this relationship. The inter-entity relationship in the dimension of Guanzhong residents' cultural digital resources needs to be analyzed from the two modes of graphic and text. The textual modal entity relationship mainly contains the relationship between the content dimension entities of Guanzhong residents' cultural projects extracted from the Guanzhong residents' cultural digital resources corpus. Using the image similarity calculation algorithm, we get the similarity relationship between the overall visual entities in the image modality, the inclusion relationship between the overall visual entities and the target visual entities through the target detection algorithm, the target visual entities then get the co-occurrence association relationship through the association rule mining, and we get the graphic linking relationship between the overall visual entities and the Guanzhong Residents' Cultural Program content dimension entities using the graphic knowledge linking method. The graphic knowledge linking method is used to get the graphic linking relationship between the overall visual entities and the content dimension entities of Guanzhong residents' cultural programs.

The attribute type analysis of Guanzhong residents' cultural multimodal resources is a further addition to the modal layer design of Guanzhong residents' cultural multimodal resources. Entity attributes of Guanzhong residents' cultural project content dimension include project attributes such as project number and project name, character attributes such as inheritor and protection unit, material attributes such as representative works and research results, and spatial-temporal attributes such as project release time and reporting area. The entity attributes of the modal dimension of Guanzhong residents' cultural projects include modal type, text and image feature attributes, and so on. Using the above features, we can further calculate the degree of correlation between the entities of Guanzhong residents' cultural digital resources, and provide support for multimodal knowledge graph entity disambiguation and knowledge inference.

This study expands the scope of Guanzhong residents' cultural digital resource organization in the modal dimension of Guanzhong residents' cultural items, and supplements the image modal knowledge types and graphic-text multimodal relationships, which makes the modal types more diversified, the entity granularity more specific, and the graphic-text relationships more rich, and provides a good basis for high-quality realization of the top-down multimodal knowledge mapping of Guanzhong residents' culture. This lays the foundation for the construction of top-down multimodal knowledge map of Guanzhong residents' culture.

(2) Entity Recognition Model

The Bert-wwm-ext-BiLSTM-Attention-CRF model proposed in this paper is based on the improvement of the Bert-BiLSTM-CRF model, which has a better performance in the field of entity recognition, and the model workflow is as follows:

Firstly, initialize the corpus and word vector training by Bert-BiLSTM-ext, use the pre-training model to encode the input graphic data of Guanzhong residents' cultural domain accordingly, extract the feature representation of the graphic, and then generate the word vector at the character level, and use the output location information and features as inputs to BiLSTM, and after that, use the BiLSTM layer to capture the order in the graphic data features to help the model better understand the temporal order information in the graphic data, learn the context information effectively in a bidirectional way, obtain the global features to get the joint vector sequences, and the output has both long and short-term memory, then process the vector sequences output from the BiLSTM layer through the attention mechanism to increase the weight for the key words in the current sequence, which can better obtain the local features, and finally, the semantic features after reinforcement by the attention mechanism are inputted into the CRF layer to ensure the reasonableness of the output labels, and the global optimization of the output label sequence to get the optimal Guanzhong residents' cultural entity recognition effect.

The structure of the Bert-wwm-ext-BiLSTM-Attention-CRF model is shown in Figure 4. The model is divided into 4 parts.

1) Bert-wwm-ext layer

Bert-wwm-ext is a pre-trained language model which is a variant of Bert. It uses the WWM approach where the whole word is masked as a unit during pre-training. This approach allows for better retention of lexical information and improves the performance of the model on tasks such as named entity recognition.

2) BiLSTM layer

BiLSTM is an RNN structure for sequence annotation task for further encoding and feature extraction on top of Bert-wwm-ext output.BiLSTM is able to capture bi-directional contextual information and model the input sequences, providing richer semantic information and contextual representation. In the named entity recognition of Guanzhong residents' cultural dataset, BiLSTM additionally encodes the Bert-wwm-ext output to extract more representative entity features. The forward and backward LSTM can be obtained from the input $x_1 \sim x_n$ for the forward output representation $\overrightarrow{h_1} \sim \overrightarrow{h_n}$ and the backward output representation $\overleftarrow{h_n} \sim \overleftarrow{h_1}$, respectively.

3) Attention layer

The role of the attention mechanism in the model is to provide better encoding of sequence information and attention to critical information. It enhances the model's attention to the important parts of the sequence by weighting the results of BiLSTM encoding and improves the model's understanding and representation learning of the input sequence.

4) CRF layer

CRF is a statistical model commonly used in sequence annotation tasks, which can globally optimize the sequences based on the constraints between the labels by taking into account the dependencies between the labeled sequences through the transfer matrix in the CRF layer. It makes the final predicted label sequences more consistent with the actual named entity boundaries and constraints.CRF takes the sequences from the Attention layer and gets the label transfer matrix by training to get the label sequence $C = \{c_1, c_2, \cdots, c_n\}$.

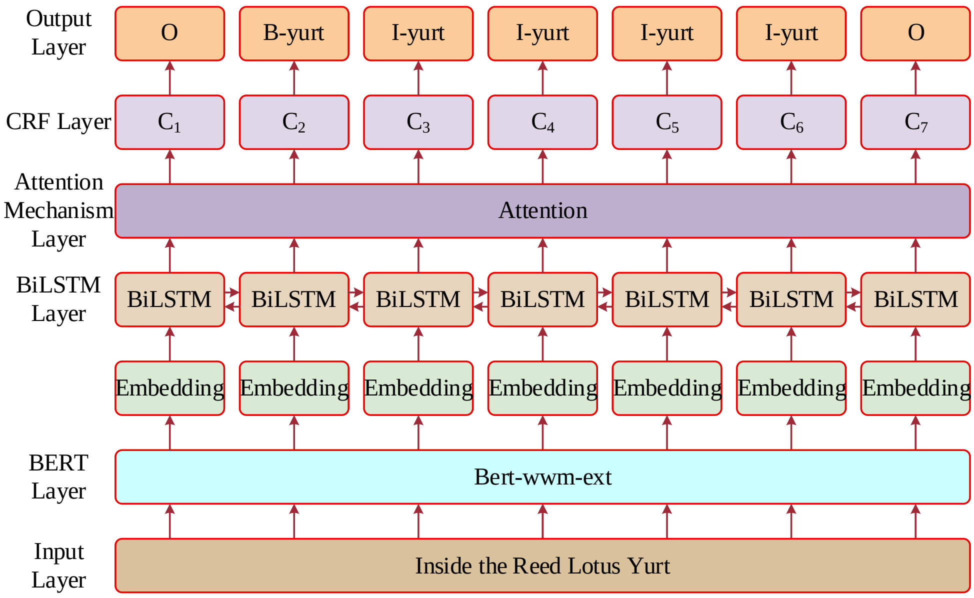

Figure 4: Bert-wwm-ext-BiLSTM-Attention-CRF model structure

(3) Relational extraction model

The BiLSTM-PCNN-Attention model proposed in this paper is based on the improvement of the PCNN-Attention model which currently performs well in the field of relation extraction, and the model workflow is as follows:

Firstly, the sentence is pre-trained and processed to obtain word vectors and position vectors, and then the vectors are used as inputs to the BiLSTM layer. Using the bidirectional LSTM algorithm, the information of context and the relationship between sentences are extracted to get the relevant graphic features and sequence features, and then the similar sentences appearing at the same time of Guanzhong residents' cultural entities are given higher weight information through the attention mechanism, and finally they are inputted to the PCNN layer, which extracts and learns the local features of the sentence information through the segmentation pooling method, extracts the optimal feature representation, and finally calculate the probability of the relationship between each entity pair by Softmax function and give the final prediction.

The BiLSTM-PCNN-Attention model architecture is shown in Figure 5. The model is divided into five parts.

1) Vector mapping layer

In the relational extraction of Guanzhong residents' cultural domain, the BERT-wwm-ext model is chosen to replace the word embedding model, and the word vectors and position vectors generated by the BERT-wwm-ext model are used as inputs to BiLSTM to obtain a richer and more accurate textual representation.

2) BiLSTM layer

In the BiLSTM layer, this paper takes the output of the vector mapping layer as the input of the BiLSTM network, and extracts the information of the context and the relationship between sentences through the bidirectional LSTM algorithm to get the relevant semantic features and sequence features.

3) Attention mechanism layer

In the attention mechanism layer, the main role is to enable the model to automatically focus on the part of the text that is relevant to the task. When the feature representation of the text is obtained from the BiLSTM layer, the attention mechanism layer evaluates the importance of each part of these features for the final task and assigns different weights to improve the accuracy of the relationship extraction.

4) PCNN layer

At the PCNN layer, this paper extracts and learns the local features of sentence information by segment pooling to extract the optimal feature representation.The PCNN model divides the input sequence into multiple segments according to the positions of the entity pairs, and then applies a convolutional neural network to each segment for feature extraction.The position sensitivity of the PCNN allows the model to better understand the position of the entities in the text and to extract the features that are conducive to the distinguishing features that distinguish the relationships between entities, thus better recognizing the relationships between cultural entities of Guanzhong residents. The input vector is partitioned into three segments as $C=\{c_{i1},c_{i2},c_{i3}\}$, and the final pooled vector is represented as $P=\{p_{i1},p_{i2},\cdots,p_{in}\}$.

5) Softmax layer

The last layer of the BiLSTM-PCNN-Attention model is the Softmax function, which mainly performs the relationship classification, calculates the probability of the relationship between each entity pair, and realizes the prediction of the relationship classification, and the calculation process is shown in Eqn. (9), which adds the result of the calculation of the upper layer, s, to the current weighting vector, z, and selects the relationship with the largest probability value as the the final prediction result. Where L is the number of label types and j is the current predicted label:

$$p(y=j|s)=\frac{e^{s^{v}z_j}}{\sum_{l=1}^{L}e^{s^{v}z_l}} \tag{9}$$

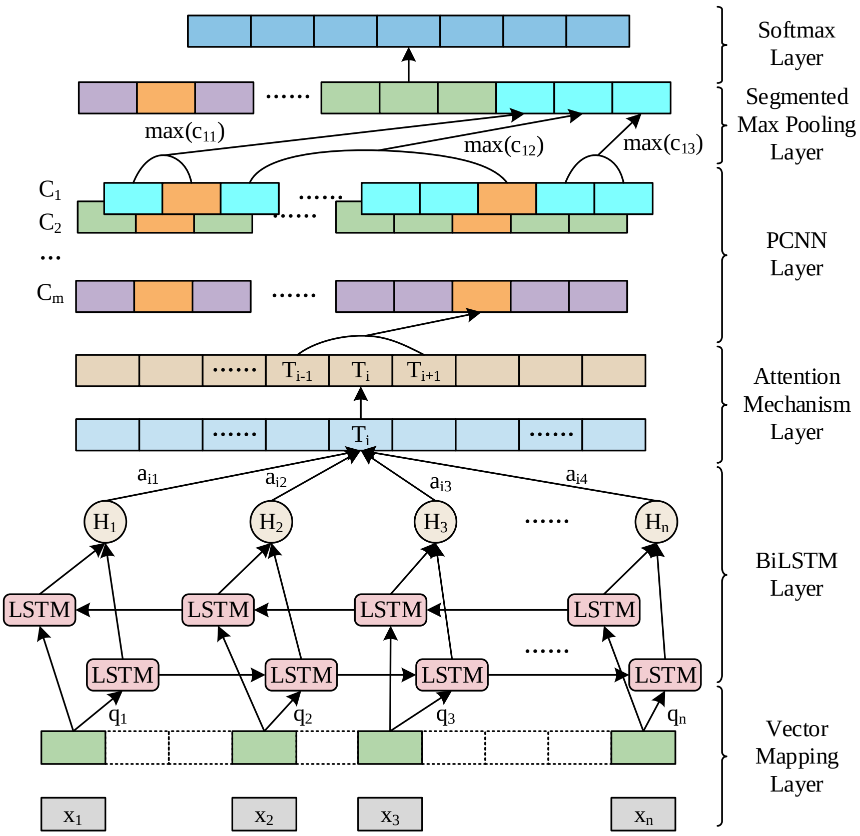

Figure 5: BiLSTM-PCNN-Attention model structure

3 Validation of the construction method of Guanzhong folk house culture gene knowledge map

In this paper, we design to use Bert-wwm-ext-BiLSTM-Attention-CRF and BiLSTM-PCNN-Attention to construct Guanzhong residential culture gene knowledge graph, in order to validate the feasibility of the knowledge graph construction method, this chapter designs the experiments to evaluate the performance of the two models in entity recognition and relationship extraction respectively.

3.1 Data sets

Two datasets, Guanzhong-MER Dataset and CLUE-Guanzhong, are chosen to validate the effectiveness of the Guanzhong folk culture gene knowledge map construction method.

Guanzhong-MER Dataset contains about 5,000 high-definition images of Guanzhong residents, folk artifacts, and festival scenes, and the amount of text data is about 5,000, which corresponds to the image data one by one.CLUE-Guanzhong contains about 10,000 Guanzhong culture-related cultural fragments, and about 2,000 pairs of Guanzhong culture scene images.

3.2 Evaluation indicators

The Guanzhong residential culture dataset constructed in this paper is labeled by multiple people in multiple rounds. In terms of text data, each person has different semantic understanding of entities and relationships. In terms of images, due to the problem of shooting angle, there will be two components appearing at the same time, and each person's judgment angle is also different, so the results are different after multi-person multi-round labeling. Different labeling results will affect the following model learning effect and the accuracy of information extraction.

In this paper, precision rate, recall rate and F1 value are used to verify the uniformity of the labeled results, which are calculated as shown in Eq. (10), Eq. (11) and Eq. (12):

$$\Pr ecision = \frac{TP}{TP+FP}*100\% \tag{10}$$

$$\operatorname{Re} call = \frac{TP}{TP+FN}*100\% \tag{11}$$

$$F1 = 2\times\frac{P\times R}{P+R}*100\% \tag{12}$$

where TP indicates an actual positive sample with a positive prediction. FP denotes a sample that is actually negative and the predicted result is positive. FN indicates an actual positive sample with a negative prediction.

The area under the ROC curve can be expressed by the AUC value, which is an index with the ability to comprehensively evaluate the prediction model, the closer the curve is to the upper left corner, the larger the AUC value is, and the better the model's performance in distinguishing between positive and negative examples.

3.3 Experimental Analysis of Guanzhong Folk Houses Cultural Entity Recognition

3.3.1 Baseline model

In order to verify the practical effect of Bert-wwm-ext-BiLSTM-Attention-CRF in multimodal named entity recognition task, the experiment compares the models that have performed well in recent years, and further compares and analyzes the model performance.

(1) VG: The overall framework of this model is based on HBiLSTM-CRF, which uses an image attention model and a gate mechanism to mine the relevant information in the picture, which in turn aids textual representation.

(2) ACoA: The overall framework of this model is based on CNN-BiLSTM-CRF, and an adaptive collaborative attention network is designed to learn image-aware text representation and word-aware image representation.

(3) UMT: Based on the Transformer structure to fuse text and image features, and combined with the plain text label detection module to assist in confirming entity boundaries.

(4) UMGF: Based on UMT, visual guidance based on target detection replaces the original image representation and constructs a graph structure of text and image target elements, which has further effect enhancement.

3.3.2 Parameter setting

Bert-wwm-ext-BiLSTM-Attention-CRF model has a word vector dimension of 764 dimensions for its output layer. And the image feature vectors are also unified to 764 dimensions in order to be consistent with the text and do the linear transformation processing. The maximum length of the sentence is 126, and the shortfall needs to be filled in to make up for it.

3.3.3 Experimental results and analysis

The performance values of each model on Guanzhong-MER Dataset are shown in Figure 6. Whether based on plain text data or graphic data, the Bert-wwm-ext-BiLSTM-Attention-CRF model outperforms the other models in practice in the multimodal named entity recognition task, with entity recognition accuracies of 76.39% and 77.84%, respectively.

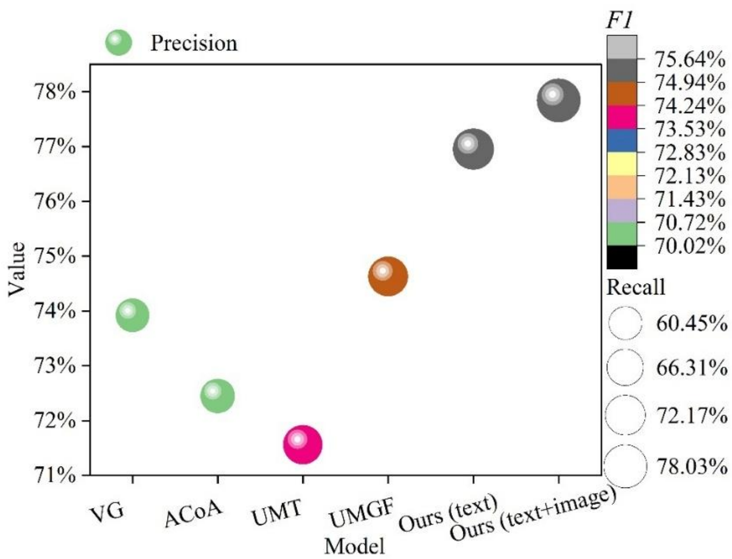

Figure 6: The performance values of the models in Guanzhong-MER Dataset

The performance values of each model on CLUE-Guanzhong are shown in Fig. 7.The Bert-wwm-ext-BiLSTM-Attention-CRF model has accuracies of 87.53% and 88.96% on text-only and graphic data, respectively, which are also better than the other compared models.

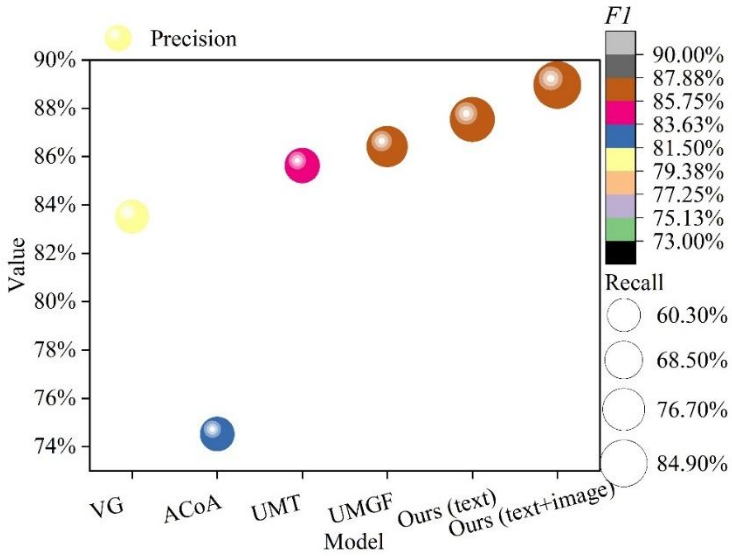

Figure 7: The performance values of the models in CLUE-Guanzhong

The Bert-wwm-ext-BiLSTM-Attention-CRF model outperforms the listed baseline models on both the Guanzhong-MER Dataset and the CLUE-Guanzhong dataset, proving that using Bert-wwm-ext-BiLSTM-Attention-CRF for modal data interaction fusion is effective. Also comparing the performance of the models on plain text data as well as graphic data, it is found that the information from multimodal fusion will be richer and have better performance than unimodal.

3.3.4 Ablation experiments

In order to further illustrate the role of multimodal interaction, this paper also conducts ablation experiments, and the experimental results are shown in Fig. 8, (a) and (b) represent the performance values of the model on the Guanzhong-MER Dataset and CLUE-Guanzhong dataset, respectively.Bert-wwm-ext-BiLSTM-Attention-CRF After removing the Bert-wwm-ext model, the model precision, recall, and F1 values are decreased, and after removing the attention mechanism, the precision, recall, and F1 values of the base model Bert-BiLSTM-CRF on Guanzhong-MER Dataset are obtained to be 71.33%, 76.84%, and 74.69%, respectively, and the precision, recall, and F1 values of the base model are obtained to be 71.33%, 76.84%, and 74.69%, respectively, and the F1 values of the base model are obtained to be 71.33%, 76.84%, and 74.69% on CLUE -Precision, recall and F1 value on Guanzhong are 85.41%, 83.22% and 84.93%, respectively. After removing the Bert-wwm-ext model and the attention mechanism, the performance of the base model is far inferior to the entity recognition model designed in this paper.

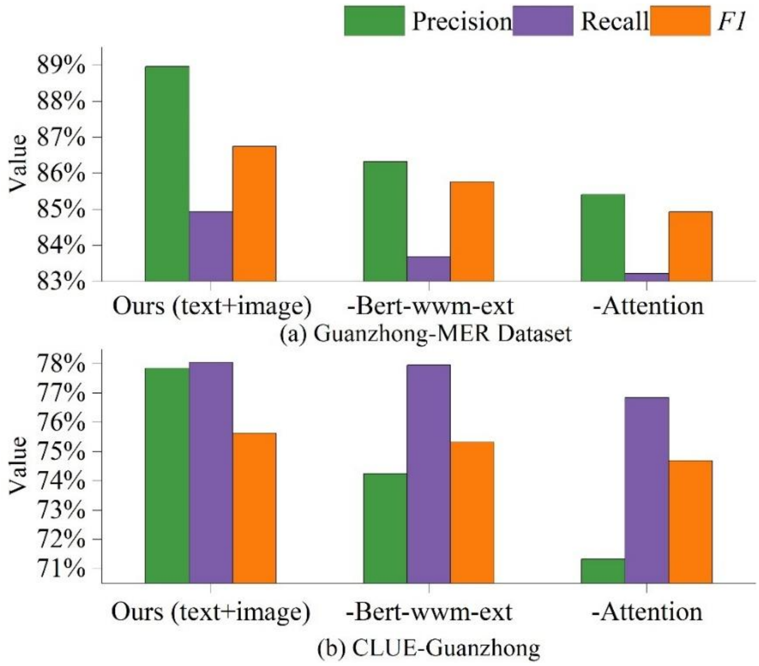

Figure 8: Ablation experiment results of Bert-wwm-ext-BiLSTM-Attention-CRF

3.4 Guanzhong Folk House Cultural Relationship Extraction Experimental Analysis

3.4.1 Baseline model

In this paper, we compare the BiLSTM-PCNN-Attention model with several character-based baseline models.

(1) BLSTM: Bidirectional LSTM is utilized to extract relations.

(2) CNN: employs one CNN result for relationship classification.

(3) PCNN: A segmented CNN model with multi-instance learning is designed.

(4) PCNN-Attention: integrating selective attention mechanism into PCNN for relation extraction.

3.4.2 Parameter setting

In this paper, we use the SGD optimizer for model training with learning rates of 0.0015 and 0.001 on the Guanzhong-MER Dataset and CLUE-Guanzhong datasets, respectively. All experiments are performed on a single NVIDIA 1080Ti GPU.

3.4.3 Experimental results and analysis

(1) Overall performance

The F1 and AUC results of each model on the two datasets are shown in Fig. 9, and (a) and (b) denote the performance values of the models on the Guanzhong-MER Dataset and CLUE-Guanzhong dataset, respectively. On Guanzhong-MER Dataset and CLUE-Guanzhong dataset, it can be seen that the BiLSTM-PCNN-Attention model can achieve the best performance, based on the plain text data, the F1 value of this paper's model reaches 51.44% and 69.85%, respectively, and based on the graphic-text interaction data, the F1 value reaches 51.33% and 69.47%, respectively. In particular, the BiLSTM-PCNN-Attention model has more obvious performance on the CLUE-Guanzhong dataset, with AUC metric values of 66.93% and 66.37% on text-only data and graphic interactive data, respectively. Compared with the PCNN-

Attention model without the introduction of BiLSTM, the F1 value of the BiLSTM-PCNN-Attention method obtains an improvement of 5.22% on the Guanzhong-MER Dataset dataset (text-only data), but only 5.11% on the graphic-text interactive data due to the fact that on multimodal relationship extraction on multimodal data, the relationship between entities is more complex, and the model performance decreases as a result.

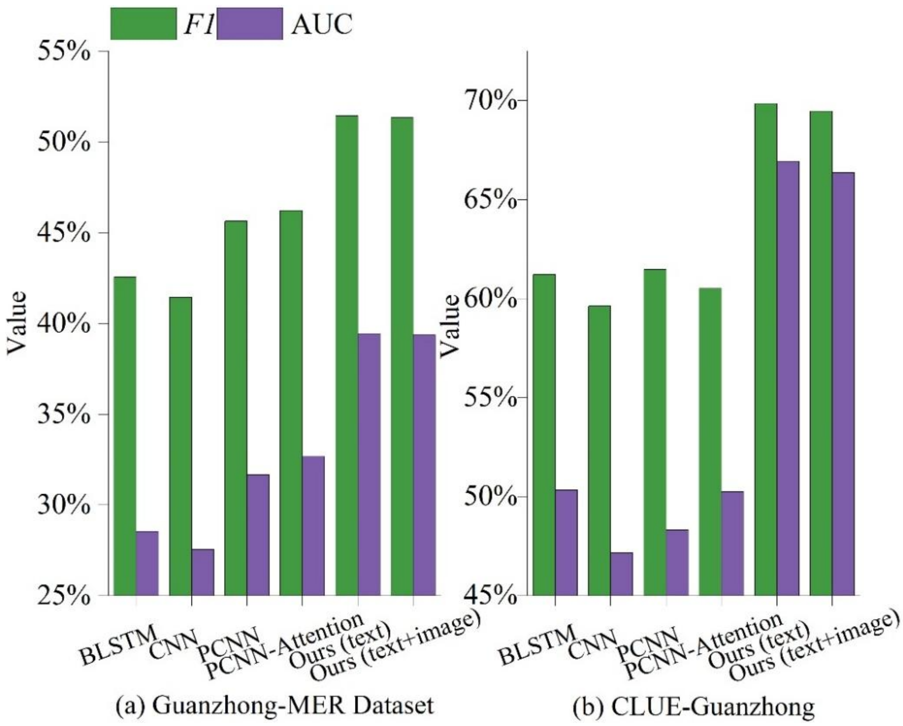

Figure 9: F1 and AUC results in two data sets

(2) Analysis of model efficiency effects

1) Analysis of the effect of different sentence lengths

In order to study the effect of different sentence length models, we analyze the performance of BiLSTM-PCNN-Attention model and PCNN-Attention baseline model on different sentence lengths on the CLUE-Guanzhong dataset, and the performance of F1 and processing speed for different sentence lengths are shown in Fig. 10, and (a) and (b) denote the F1 value and the processing speed. The sentence lengths are categorized into five groups ([0-29], [30-44], [45-59], [60-74], [≥75]). It can be observed that BILSTM-PCNN-Attention outperforms the baseline method PCNN-Attention for all different sentence lengths. In particular, the F1 value of BILSTM-PCNN-Attention is 8.55% higher than that of PCNN-Attention when the sentence length is [45-59]. In addition, we evaluated the processing speed for different sentence lengths. As can be seen in Fig. (b), BILSTM-PCNN-Attention consistently runs faster than the benchmark model for different sentence lengths. Moreover, as the sentence length increases ([30-60]), the speed difference is getting larger due to the fact that there are more sentences in the range of [30-60] than the other ranges.

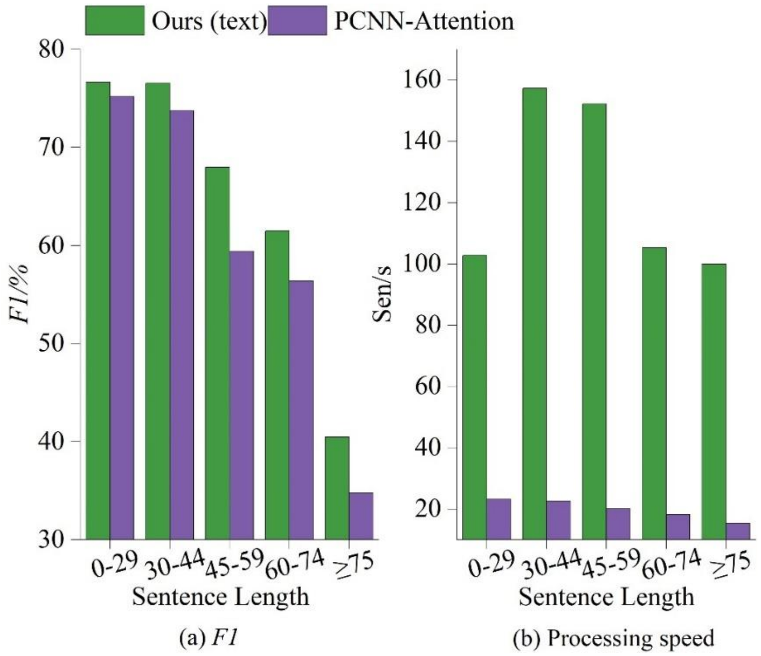

Figure 10: The speed of F1 and processing of different sentences

2) Graphic interaction effect analysis

Comparing the F1 values of this paper's model and other comparative models on the graphic interaction data, the graphic interaction effect analysis is shown in Table 2. The comparison reveals that the F1 value of this paper's model is 11.88%, 14.16%, 11.52% and 12.88% higher than that of BLSTM, CNN, PCNN, PCNN-Attention respectively under fused multimodal data. It shows that the BILSTM-PCNN-Attention model exhibits better performance on the relation extraction task.

Table 2: Graphic interaction analysis

Model	*F1*
BLSTM	0.6122
CNN	0.5963
PCNN	0.6147
PCNN-Attention	0.6052
Ours (text+image)	0.6947

3.4.4 Ablation experiments

The results of the ablation experiments of the BILSTM-PCNN-Attention model are shown in Fig. 11, and (a) and (b) denote the performance values of the model on the Guanzhong-MER Dataset and the CLUE-Guanzhong dataset, respectively.After the removal of BILSTM from BILSTM-PCNN-Attention, the model F1 value decreases to 46.22% and 60.52% on the two datasets, and after removing the attention mechanism, the F1 value decreases to 45.63% and 61.47%, respectively. It shows that after removing BILSTM and attention mechanism, the performance of the base model is far inferior to the relational extraction model designed in this paper.

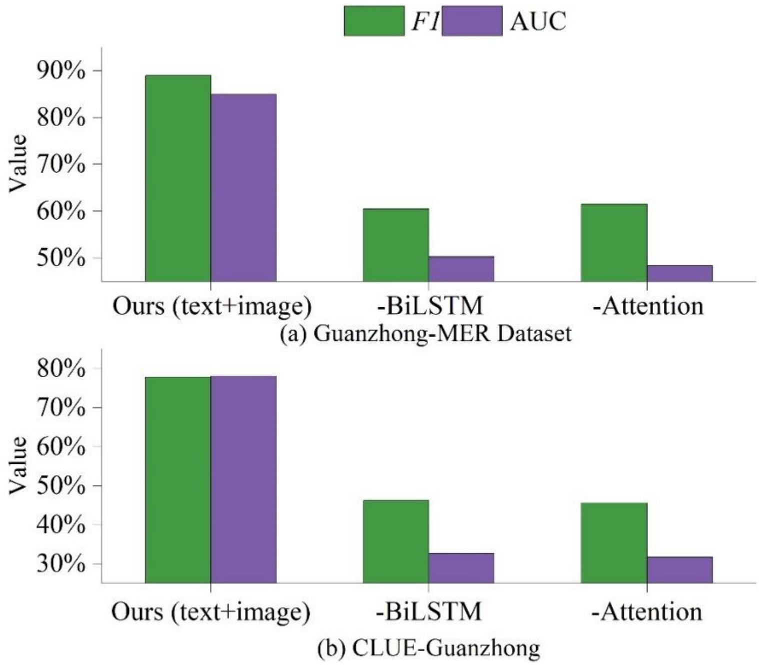

Figure 11: Ablation experiment results of BILSTM-PCNN-Attention

Through the analysis of multi-model comparison experiments, the advantages of the Bert-wwm-ext-BiLSTM-Attention-CRF and BiLSTM-PCNN-Attention models proposed in this paper over other models in terms of entity recognition and relationship extraction performance are verified, indicating that the two models proposed in this paper are more suitable for constructing the Guanzhong folk house cultural gene knowledge map.

4 Conclusion

The study improves the traditional method of constructing knowledge graph and proposes to construct Guanzhong residents' cultural gene knowledge graph based on Bert-wwm-ext-BiLSTM-Attention-CRF and BiLSTM-PCNN-Attention, and chooses Guanzhong-MER Dataset and CLUE- Guanzhong Dataset and CLUE-MER Dataset are used to test the performance of the two designed models. Bert-wwm-ext-BiLSTM-Attention-CRF has accuracies of 71.33% and 85.41% on the two datasets, and the model outperforms the base model. BiLSTM-PCNN-Attention achieves F1 values of 51.33% and 69.47% on graphic interaction data, which is more advantageous in terms of relationship extraction performance compared to PCNN and PCNN-Attention models. The experimental results show that the two models selected in this paper are able to perform entity identification and relationship extraction better in the process of cultural gene knowledge map construction, therefore, using the method of this paper to construct Guanzhong folk house cultural gene knowledge map can fully present the relationships in the field of Guanzhong folk house culture, which is conducive to the protection and development of Guanzhong folk house culture.

Funding

This work was supported by "Ridge of China - Research on Qinling Cultural Heritage and Environmental Aesthetics" 2022 Humanities and Social Sciences General Project of Ministry of Education of the People's Republic of China (22YJA760045); Shaanxi Province Youth

Innovation Team Building Scientific Research Project “Research on Chinese Traditional Construction System and Contemporary Environmental Art Design” (21JP091)；Discipline construction Project of Xi'an Academy of Fine Arts"Artistic Exchange and Cultural Integration of the Silk Road: Environmental System Regeneration and Dissemination of Local Heritage"(XK202601).

About the Author

Hao Zhang is an associate professor and master's supervisor at Xi'an Academy of Fine Arts, specializing in Architectural Environmental Art. He holds a Ph.D. in Design. Research interests: Architectural cultural heritage conservation, environmental art design, public art design. Member of the China Institute of Architects, Council Member of the Interior Design Branch of the China Institute of Architects, Deputy Secretary-General of the Environmental Art Professional Committee of the China Cultural Promotion Association, Member of the Environmental Art Committee of the Shaanxi Artists Association, Core Member of the Shaanxi Higher Education Youth Innovation Team, Council Member of the Xi'an Civil Engineering and Architecture Society.

References

[1] Li, Z., & Xu, J. S. (2014). On the abstract metaphorical design in the transformation of traditional houses in Guanzhong region. Applied Mechanics and Materials, 507, 19-30.

[2] Song, B., Bai, L., Wu, L., & Yang, L. (2022). Optimization of Energy Conservation for Rural Residential Buildings in the Guangzhong Plain. In Hydraulic and Civil Engineering Technology VII (pp. 1147-1154). IOS Press.

[3] Lian, M., & Li, Y. (2024). The Spatial Patterns and Architectural Form Characteristics of Chinese Traditional Villages: A Case Study of Guanzhong, Shaanxi Province. Sustainability (2071-1050), 16(21).

[4] Na, W., Jia, W., & Juan, X. (2015). The Evolution of Dwellings Spatial Pattern and its Influence Factors in Guanzhong Region. Open Cybernetics & Systemics Journal, 9, 1611-1617.

[5] Abd Manaf, A., Latip, N. S. A., Ismail, N. H., Ismail, W. N. W., Ab Rashid, N. S., & Nizar, W. I. W. A. (2025). Challenges in Adapting Vernacular Architecture to the Contemporary: A Systematic Literature Review. Semarak International Journal of Design, Built Environment and Sustainability, 3(1), 17-29.

[6] Guerrero Baca, L. F., & Soria López, F. J. (2018). Traditional architecture and sustainable conservation. Journal of Cultural Heritage Management and Sustainable Development, 8(2), 194-206.

[7] Hogan, A., Blomqvist, E., Cochez, M., d’Amato, C., Melo, G. D., Gutierrez, C., ... & Zimmermann, A. (2021). Knowledge graphs. ACM Computing Surveys (Csur), 54(4), 1-37.

[8] Chen, X., Jia, S., & Xiang, Y. (2020). A review: Knowledge reasoning over knowledge

graph. Expert systems with applications, 141, 112948.v

[9] Chansanam, W., Jaroenruen, Y., Kaewboonma, N., & Tuamsuk, K. (2022). Culture knowledge graph construction techniques. Education for Information, 38(3), 233-264.

[10] Li, X., Xie, R., & Xia, H. (2024). Research on the construction of knowledge map of ethnic minority cultural information resources. Advances in Education, Humanities and Social Science Research, 12(1), 50-50.

[11] Yang, S., Yi, L., Guan, H., & Li, Y. (2023). A meme-based approach for knowledge mining, organization, and presentation of traditional Chinese settlement culture. Heritage Science, 11(1), 206.

[12] Liu, B., Zhao, W., Wang, J., Yan, J., & Peng, J. (2023, May). Construction and Application of Multi-modal Knowledge Graph for Collection of Cultural Relics. In 2023 2nd International Conference on Image Processing and Media Computing (ICIPMC) (pp. 142-149). IEEE.

[13] Fan, T., Wang, H., & Hodel, T. (2023). CICHMKG: a large-scale and comprehensive Chinese intangible cultural heritage multimodal knowledge graph. Heritage Science, 11(1), 115.