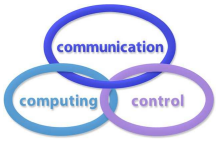
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 21, Issue: 4, Month: August, Year: 2026
Article Number: 7233, https://doi.org/10.15837/ijccc.2026.4.7233



CCC Publications

A joint-embedding framework fusing multi-feature information for cultural relic entity alignment in knowledge graphs

Min Zhang,Luya Yang,Yaxian Gao,Yong Ren

Min Zhang*
School of Information Engineering
Shaanxi Xueqian Normal University, China
No. 101, Shenhe 2nd Road, Chang'an District, Xi'an City, Shaanxi, China
*Corresponding author: 28028@snsy.edu.cn

Luya Yang
School of Information Engineering
Shaanxi Xueqian Normal University, China
No. 101, Shenhe 2nd Road, Chang'an District, Xi'an City, Shaanxi, China
28032@snsy.edu.cn

Yaxian Gao
School of Information Engineering
Shaanxi Xueqian Normal University, China
No. 101, Shenhe 2nd Road, Chang'an District, Xi'an City, Shaanxi, China
28013@snsy.edu.cn

Yong Ren
Xi'an Shiyou University, China
No. 18, Dianzi Erlu, Yanta District, Xi'an City, Shaanxi, China
renyong@xsyu.edu.cn

Abstract

Entity alignment across multi-source knowledge bases is frequently hindered by the problem of 'heterogeneous names but equivalent semantics,' a challenge that becomes particularly acute in domains with specialized nomenclature and rare characters, such as cultural relics. To address this universal issue, this paper proposes a joint-embedding framework, MulF-ELMo-BERT, suitable for entity alignment in cultural relic knowledge graphs. This framework integrates multi-dimensional features such as entity names, attributes, summaries, and full-text texts, achieving comprehensive extraction of entity features from four levels: characters, words, sentences, and paragraphs. It effectively filters out weakly relevant entities and breaks through the limitations of semantic representation relying solely on a single feature. Given the dense presence of rare characters and specialized terms in cultural relic entity names, the ELMo context-aware word embedding model is introduced. It can dynamically adjust the vector representation of rare words based on their context, significantly enhancing semantic adaptability. Meanwhile, a Chinese BERT model with an integrated whole-word masking strategy is adopted to avoid local co-occurrence interference

and strengthen the ability to capture term associations. More importantly, the framework deeply fuses ELMo embeddings with Chinese BERT model embeddings as entity embeddings, effectively compensating for the limitations of a single embedding model in complex semantic representation. By then adaptively calibrates cosine similarity using a learnable threshold parameter, the matching accuracy is improved, enabling the alignment of cultural relic entities. Experimental results demonstrate that the proposed method can efficiently capture multi-level semantic features of entities and exhibits excellent performance in cultural relic entity alignment tasks. The proposed framework presents a generalizable architecture for multi-feature entity alignment, with its effectiveness rigorously validated on the complex domain of cultural relics.

Keywords: Multi-feature fusion, entity alignment, Joint-embedding, Encyclopedic knowledge base.

1 Introduction

The integration of multi-source knowledge bases is pivotal for constructing comprehensive knowledge graphs. A fundamental challenge in this process is Entity Alignment, which aims to identify entities that refer to the same real-world object despite having different surface forms across sources. This problem is ubiquitous but manifests with unique intensity in domains characterized by specialized terminologies and heterogeneous data schemas[1]. As a compelling case in point, the cultural relic domain exemplifies these challenges[2]. For instance, the same entity, such as the renowned playwright "William Shakespeare", may be referred to by various names across different data sources (e.g., "Shakespeare," "W. Shakespeare," or "The Bard"). Therefore, entity alignment is required to identify and match these semantically identical entities that differ in their surface forms, enabling the integration and fusion of multi-source knowledge bases.

However, due to the diversity of data sources and the differences between manual input and definitions on encyclopedic websites, entities with the same semantics have different forms of representation in different knowledge bases. The alignment of cultural relic entities across multi-source knowledge bases is a crucial step in the integration of cultural relics knowledge graphs, while also posing certain challenges.

For encyclopedic knowledge bases, the existing entity alignment technologies mostly rely on the similarities of the knowledge base system or use entity information to calculate the similarities between entities of different knowledge bases to achieve the goal of knowledge base fusion. Gerlach et al.[3] proposed proposes a machine-in-the-loop entity linking system, aiming to increase the link coverage of new pages and Wiki projects with low resources. Gabri et al.[4] proposed the explicit semantic analysis method. Lin et al.[5] proposed an end-to-end entity recognition and disambiguation framework for identifying author affiliation information from literature publications.

However, the above methods mainly rely on the original page relocation disambiguation function of the encyclopedia knowledge base, only use a single feature of each encyclopedia entity, and fail to make good use of the multiple types of encyclopedia knowledge information.

To alleviate the limitation of single-entity features in the entity alignment task, many researches fuse multiple features of an entity for entity alignment. Zhang et al.[6] proposed a selective filtering entity alignment method that improves entity alignment through attribute filtering and external knowledge filtering. Yang et al.[7] proposed a pseudo-connected network, which adopts the common concern method to collaboratively learn attribute types and attribute values.

However, the entity alignment task considers multiple features such as entity attributes and merely acquires the similarity between two entities at the level of characters and words. It does not consider the meaning of words or the relationship between words, nor does it take into account the context information between sentences.

In entity alignment, contextual information is crucial for determining whether two entities refer to the same real-world object. The same entity may exhibit different manifestations or meanings in various contexts[8]. word embeddings can capture contextual information, enabling more accurate judgments of entity similarity and thereby enhancing the precision of entity alignment. Therefore, researchers focusing on entity alignment attach great importance to the study of word embeddings [9**?**]. Yu et al.[11] proposed a pre-trained joint alignment method based on BERT and GloVe word embed-

dings. The joint alignment of attributes and relations using multiple pre-trained models can improve the model performance. Sun et al.[12] proposed a two-stage strategy for aligning both homogeneous and heterogeneous entities. In the first stage, an embedding-based semantic clustering algorithm is employed to partition the semantic space into multiple clusters, and the clusters are then paired based on centroid distances. Zhong et al.[13] proposed a semantic-driven entity embedding method. This method comprises attribute embedding and relation embedding. Specifically, the attribute embedding utilizes a pre-trained transformer-based language model to capture semantic information from attribute values.

However, in the field of cultural relics, the complexity of word formation such as rare characters and professional terms in entity names makes it difficult for the existing models to handle these unregistered or sparse words through word embedding. Moreover, the existing models only generate word vectors based on the co-occurrence statistics of words, ignoring the context information, resulting in the inability to accurately capture the semantics of entities in the entity alignment task.

In recent years, deep models have been widely applied in entity alignment tasks. Especially when pre-trained language models are trained on large-scale corpora, they can learn rich language knowledge and semantic representations. Ge et al.[14] proposed an efficient name-based entity alignment process. It eliminates preprocessing through the name information provided by the entity itself and combines and fuses the features captured by the entity name and the structural information of the graph to improve the entity alignment result. Liu et al.[15] proposed to utilize the Graph Neural Network (GNN) attribute value encoder and divide the Knowledge Graph (KG) into subgraphs to effectively model various types of attribute triples. Paganelli et al.[16] conducted multi-faceted analyses using components of the pre-trained and fine-tuned BERT architecture. BERT is used to generate highly contextualized term embeddings.

Existing research has made certain progress in the field of entity alignment. However, during the pre-training process of the existing models, some labels in the input are randomly masked, making it difficult to effectively capture the semantic associations and context information between terms, which reduces the generalization ability of the models.

Based on the characteristics of encyclopedic knowledge bases and the manifestation forms of cultural relic domain data within them, and targeting the multi-source heterogeneous data from Baidu Encyclopedia and Chinese Wikipedia, this paper proposes a joint embedding framework for cultural relic knowledge graph entity alignment, named MulF-ELMo-BERT. This framework integrates multi-dimensional features such as entity attributes, summaries, and full-text texts, achieving comprehensive extraction of entity features from four levels: characters, words, sentences, and paragraphs. It can effectively filter out weakly relevant entities and reduce the limitations of relying solely on single entity features from multi-source heterogeneous data in entity alignment tasks.

In terms of model design, considering the prevalence of rare characters and dense specialized terms in the names of cultural relic entities, the Embeddings from Language Models (ELMO) context-aware word embedding model is introduced. It can dynamically adjust the vector representation of rare words based on their specific contexts, significantly enhancing semantic adaptability. Simultaneously, a Bidirectional Encoder Representations from Transformers (BERT) model with an integrated whole-word masking strategy during pre-training is employed to avoid local co-occurrence interference and strengthen the ability to capture term associations.

More importantly, this framework deeply fuses ELMo embeddings with Chinese BERT model embeddings as entity embeddings. The dynamic vector adjustment capability of ELMo endows entity features with fine-grained context awareness advantages, while the whole-word semantic capture ability of Chinese BERT reinforces term-level semantic association expression. The complementary fusion of the two effectively compensates for the limitations of a single embedding model in complex semantic representation.

On this basis, a dynamic threshold determination method is used to adaptively calibrate the cosine similarity calculation results, further improving the accuracy of entity matching determination and ultimately achieving cultural relic entity alignment. Experimental results demonstrate that the proposed method can efficiently capture multi-level entity semantic features and exhibits excellent performance in cultural relic entity alignment tasks, providing a feasible solution for entity alignment

in cultural relic domain knowledge fusion and offering a reference for entity alignment research in related fields.

The primary contributions of this study are as follows:

(1) This paper proposes a multi-dimensional feature fusion-based entity feature extraction mechanism.To tackle the constraints of depending solely on single entity features from multi-source heterogeneous data, this mechanism integrates features including attributes, summaries, and full-text content. It extracts features across four levels, spanning from characters to paragraphs, filters out weakly relevant entities, breaks through traditional limitations, and offers more robust feature support.

(2) This paper puts forward a complementary fusion embedding strategy that combines ELMo and Chinese BERT. In light of the characteristics of entity names, this strategy fuses ELMo with a whole-word masking Chinese BERT model. ELMo dynamically adjusts its vector representation based on the meanings of rare words in a specific context, enhancing the semantic adaptability of rare words. Meanwhile, the whole-word masking Chinese BERT model strengthens the capture of term associations, thereby improving the semantic representation of entity embeddings.

(3) This paper introduces a dynamic threshold-assisted entity alignment determination method. Based on the fused embeddings of ELMo and BERT, this method calibrates cosine similarity using dynamic thresholds. It adjusts the determination criteria according to entity features, enhances matching accuracy, and resolves the adaptability issues associated with fixed thresholds.

2 Related Work

2.1 Entity alignment of the encyclopedia knowledge base

The differences between manual entry and definition on encyclopedic websites make the existence of synonymous and synonym entities face the problem of entity alignment in multi-source knowledge bases in the process of achieving knowledge fusion. Many scholars have conducted long-term research.

Huang et al.[17] constructed two knowledge bases, one of the Baidu Encyclopedia and one of an interactive encyclopedia, and calculated attribute similarity with the Longest Common Subsequence (LCS) algorithm to determine whether two entities link to the same object in the real world. Zhang et al.[18] applied LDA to nouns, verbs, and other words of text, respectively, to obtain the similarity of different parts of speech words and calculated the context similarity by comprehensively weighting the three similarity degrees. This method reduced time complexity due to the parallelization of the modeling process. Zhang et al.[6] proposed a selective filtering entity alignment framework, which filters the candidate set by using attribute information and external knowledge respectively, thereby deleting most of the erroneous entities in the candidate set. This framework mainly uses the selective filtering mechanism and external knowledge for entity alignment. Deng et al.[19] proposed an event entity alignment method based on event elements, and this method calculates the entity similarity based on multiple event elements. This method optimizes the threshold setting, thereby enhancing the ability to identify the existence of aligned entities. The performance of this method in terms of event entity alignment has been significantly improved.

However, the above-mentioned entity alignment task only acquires the similarity of two entities at the character and word levels, without considering the meaning of words or the relationship between words, nor taking into account the context information between sentences.

2.2 Embedding-based entity alignment

TransE is the first and also one of the most popular models proposed for the entity alignment task[20]. TransE interprets a relation as a translation operating on the low-dimensional embeddings of entities. More specifically, given a relational triple (h, r, t), it suggests that the embedding of the tail entity t should be close to the embedding of the head entity h plus the embedding of the relation r, that is, $\tilde{h} + \tilde{r} = \tilde{t}$. As a result, the structural information of the entities can be preserved, and entities sharing similar neighbors have close representations in the embedding space.

Lin et al.[5] proposed an end-to-end entity recognition and disambiguation framework. The proposed algorithm, which combines word embedding and spatial embedding, leverages semantic and

geographic information to effectively improve the performance of entity recognition and disambiguation. Yoon et al.[21] proposed an unsupervised entity alignment method, which jointly performs entity-level and relational-level alignment by using the neighbor triple matching strategy of the semantic text features of relations and entities. The accuracy of double alignment was evaluated using the semantic feature GloVe and the lexical feature Bigram. Li et al.[11]adopted joint alignment using pre-trained BERT and GloVe embeddings to leverage their complementary strengths. By integrating BERT' s contextualized representations with GloVe' s global word co-occurrence statistics, the approach captures both semantic nuances and syntactic regularities, thereby enhancing the robustness of entity matching across heterogeneous knowledge sources. Kolyvakis et al.[22] proposed an entity alignment method called DeepAlignment. This method refines the pre-trained word vectors with the aim of obtaining the description of the ontology entity suitable for the ontology matching task. A set of word2vec vectors trained on Google News data is used to verify the impact of the initial pre-trained word vectors on the performance of DeepAlignment. Shahbazi et al.[23] proposed a new local entity disambiguation system. The system has learned the entity-aware extension of Language Model Embedding (ELMo) which is called Entity-ElMO (E-ELMo). The system utilizes E-ELMo for local entity disambiguation.

Existing embedding-based entity alignment models have made certain progress in general domains. However, due to the unique nature of knowledge fusion in the cultural relic field, these existing methods fail to take into account the dense rare words and specialized terms present in relic names and lack the comprehensive utilization of multi-dimensional features. Consequently, their performance in cultural relic entity alignment may be rather poor.

2.3 Entity alignment based on deep models

Yang et al.[7] proposed a network to make fulluse of the attribute types and values in KGs. method combines the structure and attribute information of entities respectively by jointly training two embedding learning components. Use the joint attention method to collaboratively learn attribute types and attribute values. Tam et al.[9] proposed an end-to-end, unsupervised entity alignment framework for cross-language knowledge graphs. This model uses a multi-order GCN model to capture relation-based correlations between entities, and simultaneously combines attribute-based correlations through a translator. The model adopts a post-fusion mechanism to combine all the information together, thereby enhancing the final alignment result and making the model more robust. Wu et al.[24] proposed a relations-aware double-graph convolutional network (RDGCN), which integrates relational information through the close interaction between knowledge graphs and their dual relations. RDGCN learns better entity representations by capturing adjacent structures. Zhu et al.[25] proposed a relationship-aware neighborhood matching model RNM for entity alignment. The model utilizes neighborhood matching to enhance entity alignment. In addition to comparing neighboring nodes when matching the neighborhood, the model attempts to mine useful information from the connection relationship. Bai et al.[26] proposed a new entity alignment method based on BERT, which uses a multilingual trained BERT model to learn the semantic relevance of entity descriptions.

Existing deep models have made considerable progress in the field of entity alignment. However, due to the characteristics of knowledge fusion in the encyclopedia knowledge base and the particularity of entity word formation in the field of cultural relics, the existing models have difficulty effectively capturing the semantic associations and context information between terms, thereby reducing the generalization ability of the models.

Unlike frameworks like DeepAlignment that primarily refine pre-trained word vectors for similarity calculation, our approach first deeply fuses contextual embeddings from ELMo and BERT. It then introduces a learnable, dynamically tuned threshold mechanism that is jointly optimized with the embeddings. More importantly, our entire pipeline—from feature fusion to alignment decision—is integrated into an end-to-end differentiable framework trained with a unified loss function, eliminating reliance on independent, non-optimized components.

3 Method

This paper proposes a general joint-embedding framework for multi-feature entity alignment. To validate its effectiveness in complex real-world scenarios, we instantiate and evaluate the framework using cultural relic dat-a domain characterized by professional terminology and the critical challenge of heterogeneous names but equivalent semantics. The core innovation of the proposed MulF-ELMo-BERT framework lies in its end-to-end trainable paradigm, which jointly optimizes the multi-feature embedding fusion process and the entity alignment decision threshold through a unified loss function. As shown in Figure 1, the framework consists of three core modules:

(1) Data Sources and Feature Extraction.Integrates Baidu Baike and Wikipedia knowledge bases, simultaneously extracting entity names, attributes, summaries, and full-text documents. This enables semantic capture at four levels: characters, words, sentences, and paragraphs.

(2) Embedding Representation Layer. Employs the context-aware ELMo dynamic word embedding model to handle rare characters and specialized terms in the cultural relic domain. By integrating a whole-word masking strategy, it enhances the Chinese BERT's capability to capture term associations. The joint embedding of ELMo and BERT compensates for the limitations of single-model semantic representation.

(3) Entity Alignment Module. Combines local and global representations to construct feature vectors for entity pairs. Cosine similarity is used to measure semantic relevance between entities, while adaptive threshold calibration achieves precise alignment of cultural relic entities.

The framework first extracts multi-level features from multi-source knowledge bases, then obtains enhanced semantic representations through the joint embedding model, and finally completes entity alignment using optimized similarity calculation. This effectively resolves the semantic heterogeneity problem in knowledge fusion.

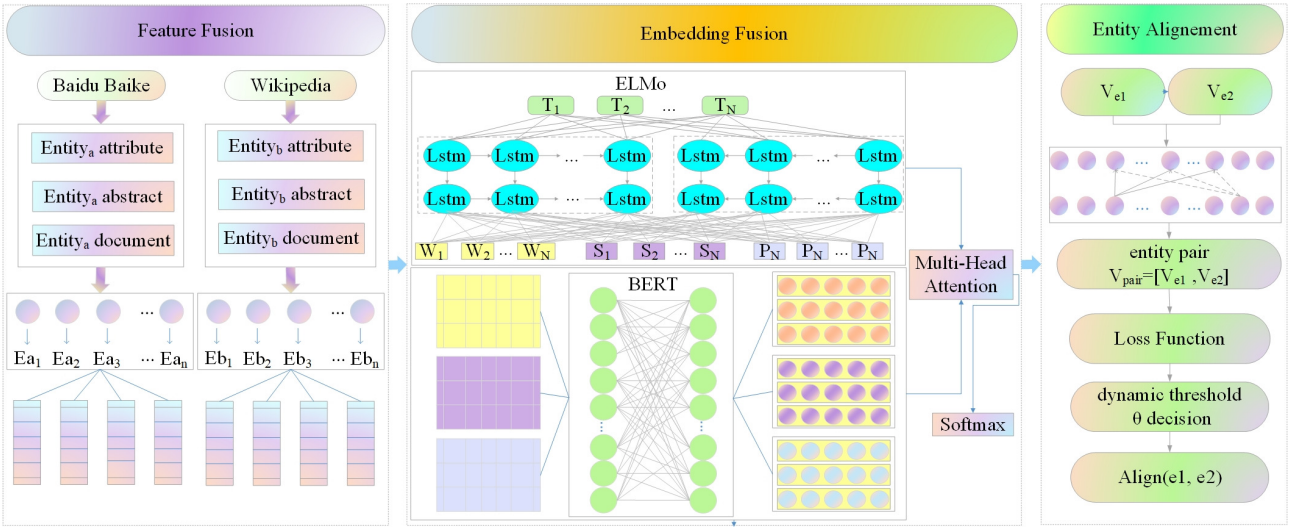


Figure 1: The framework of the MulF-ELMo-BERT model

3.1 Multi-feature Fusion

Firstly, the cultural relic data obtained from encyclopedic websites are preprocessed, this involves deleting irrelevant information, stopping words, clauses, word segmentation, part-of-speech tagging, and syntactic analysis.

After data preprocessing, they are stored as a triple $< entityattribute, entityabstract, entitycontext >$, and entities in two encyclopedic knowledge bases are formalized:$e_a =< attr, abst, docu >$ and $e_b =< attr, abst, docu >$, where e_a and e_b are the names of the entity related to cultural relics, $attr$ represents the entity attribute of cultural relics, $abst$ represents the entity abstract of cultural relics, and $docu$ represents the entity context information of cultural relics. For a cultural relic entity e_a, there are multiple attributes $attr_{ea}$, each of which has a corresponding attribute value $value_{ea}$. The entity

abstract *abst* is stored as text and forms the main content of entity description text. Entity context *docu* is the unstructured text describing the entity on the web page, also known as entity context.

The entity attributes, entity abstractions and entity contexts after word segmentation are respectively taken as the inputs of the models ELMo and Chinese-BERT, and the context feature extraction is implemented by the two models. Then, the similarities of entity attributes, entity abstractions and entity context features are fused to generate a multi-feature similarity fusion model. By comparing the calculation results of multi-feature similarity with the threshold, entity alignment is achieved. Finally, the set of entity alignment is output.

3.2 ELMo for Dynamic Contextual Representation

To address the challenge of rare characters and specialized terms in cultural relic names, whose meanings can shift depending on context, we employ the ELMo model. Its key advantage lies in generating dynamic, context-aware word embeddings. Unlike static embeddings, ELMo provides a representation for each word that is a function of the entire input sentence, enabling it to disambiguate polysemous words effectively.

ELMo computes a layered representation for each token t_i through a bidirectional LSTM, combining the outputs from all layers into a comprehensive set of representations R_i, as defined in Equation 1:

$$R_i = \{x_i^{LM}, \overrightarrow{h}_{i,j}^{LM}, \overleftarrow{h}_{i,j}^{LM} | j = 1, ..., L\} = \{h_{i,j}^{LM} | j = 0, ..., L\} \tag{1}$$

where $h_{i,j}^{LM}$ represents the character-level encoding, and $\overrightarrow{h}_{i,j}^{LM}$, $\overleftarrow{h}_{i,j}^{LM}$ denote the forward and backward hidden states at the j-th layer, respectively.

In our framework, we adapt ELMo to the entity alignment task by fine-tuning the pre-trained model and learning task-specific weights for each layer's representation. The final ELMo embedding for entity alignment, $ELMo_i^{EA}$ is a weighted combination of these layered representations, calibrated by a scaling factor γ^{EA}, as shown in Equation 2:

$$ELMo_i^{EA} = E(R_i; \theta^{EA}) = \gamma^{EA} \sum_{j=0}^{l} s_j^{EA} h_{i,j}^{LM} \tag{2}$$

Here, s^{EA} represents the softmax-normalized weights optimized for the entity alignment task.

The tokenized sequences from entity names, attributes, summaries, and full-text contexts are processed by this adapted ELMo model. The resulting contextualized vectors capture nuanced semantic and syntactic features crucial for distinguishing entities with heterogeneous names but isomorphic semantics, forming the dynamic, context-sensitive stream of input for our subsequent fusion module.

3.3 Chinese-BERT for Term-Integrated Semantic Representation

To complement ELMo's dynamic contextual capabilities and enhance the modeling of complete semantic units prevalent in the cultural relic domain, we incorporate a Chinese-BERT model pre-trained with Whole Word Masking (WWM) strategy. This approach is particularly crucial for handling multi-character professional terms like "Tang Caifengmingqi Seven-Stringed Zither", where preserving semantic integrity is essential.

To address the out-of-vocabulary (OOV) challenge posed by rare characters and specialized terms, we construct the input embedding through a hybrid character-word approach. The embedding E_i for each token is formed by summing three components, as defined in Equation 3:

$$E_i = E_i^{token} + E_i^{segment} + E_i^{position} \tag{3}$$

Notably, we employ Rotary Position Embedding (RoPE) for the positional encoding to better capture long-range dependencies in cultural relic descriptions, as shown in Equation 4:

$$\mathrm{RoPE}(q_m, k_n) = \mathrm{ReLU}(W_q q_m + W_k k_n) \cos\theta(m - n) \tag{4}$$

where $\theta(m - n)$ represents the angular encoding for the positional difference between characters.

The core innovation lies in our adaptation of the pre-training objective. We optimize the Masked Language Model (MLM) task using the WWM strategy, where the loss function is defined in Equation 5:

$$\mathcal{L}_{\text{MLM}} = -\frac{1}{N}\sum_{i=1}^{N}\log P(w_i|h_{[\text{MASK}]_i}) \tag{5}$$

Unlike character-level masking, our WWM strategy uniformly masks all characters within the same word (e.g., replacing "bronze ware" with "[MASK][MASK][MASK]"), forcing the model to reconstruct complete terms based on contextual understanding. This significantly enhances semantic modeling for cultural relic terminology. Additionally, we employ domain-specific data augmentation using the Encyclopedia of Chinese Cultural Relics, including synonym replacement and irrelevant term insertion, to further improve the model's robustness and domain adaptability.

The Chinese-BERT component provides a robust, statically-oriented semantic representation that emphasizes term integrity and global contextual associations within the text. Serving as the complementary stream to ELMo's dynamic embeddings in our fusion module, it enables a comprehensive semantic understanding of cultural relic entities.

3.4 Word Embedding Fusion Based on Multi-Head Attention Mechanism

The ELMo model captures dynamic contextual word meanings through a bidirectional LSTM structure, excelling at handling polysemy . In contrast, Chinese-BERT, based on the Transformer's self-attention mechanism, is capable of modeling long-range dependencies and reinforces the integrity of specialized terms through the Whole Word Masking (WWM) strategy. However, a single model may suffer from issues such as local feature overfitting or sparse global semantics.

To effectively integrate the complementary strengths of ELMo and Chinese-BERT in word embeddings for cultural relic entity alignment tasks, this paper proposes a dynamic word embedding fusion strategy based on the Multi-Head Attention (MHA) mechanism. This strategy captures the correlations between different models in multi-level semantic representations, enabling adaptive fusion of cross-model features and thereby enhancing the semantic representation capability for complex cultural relic terms.

We propose a multi-headed attention mechanism that integrates multi-model word embeddings to realize the joint word embedding representation of ELMo and BERT-Chinese，with specific steps as follows:

3.4.1 Feature Extraction Layer

After tokenizing cultural relic entity names, attributes, and contextual text, the pre-trained ELMo model processes them to generate dynamic word embeddings $E_i \in R^{d_e}(d_e = 1024)$, incorporating character-level, syntactic, and semantic contextual information.

The same text is fed into the Chinese-BERT model, and its final hidden layer states are extracted as static word embeddings $B_i \in R^{d_b}(d_b = 768)$, integrating character, lexical, and entity relation knowledge.

3.4.2 Multi-Head Attention Interaction Layer

To align the semantic spaces of ELMo and BERT, a cross-model attention module is designed. Using ELMo features E as the Query and BERT features B as the Key and Value, attention weights are computed by scaling the dot product attention as follows in Equation 6:

$$Attn(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{6}$$

In the formula,$Q = EW_q$, $K = BW_k$，$V = BW_v$，where W_q，W_k，W_v are learnable parameters respectively, and d_k is the dimensional scaling factor. Through multi-head parallelization (with h=8), the model can simultaneously focus on the interaction information of different semantic subspaces such

as glyphs, word meanings, and entity relations. Finally, the outputs of each head are concatenated and linearly transformed to obtain the fused feature $F_i \in R^{d_j}(d_j = 1024)$.

3.4.3 Dynamic Gating Fusion Layer

To balance the dynamic nature of ELMo and the global nature of BERT, a gating mechanism is introduced to adaptively adjust feature weights, in Equation 7 and 8:

$$G_i = \sigma(W_g[E_i; b_i]) + b_g \tag{7}$$

$$O_i = G_i \odot E_i + (1 - G_i) \odot B_i \tag{8}$$

Here, σ represents the Sigmoid function, $W_g \in R^{d_e \times 2d_e}$is the weight matrix, and $\odot$ denotes element-wise multiplication. The gating value G_i is dynamically generated based on the input features, enabling the model to rely more on ELMo's contextual adjustment capability when handling polysemous words, while focusing on BERT's complete semantic representation when dealing with specialized terms.

ELMo's dynamic word vectors capture contextual ambiguities in cultural relic terminology, while BERT's static embeddings preserve the integrity of specialized terms. These two are aligned through the multi-head attention mechanism. The hybrid embeddings and gating mechanism enhance robustness. By fusing features from both sources, the model integrates global and local information, thereby improving its ability to model long-range dependencies.

3.5 Entity Alignment

3.5.1 Entity Representation Extraction

This module proposes a dual-channel context-aware representation extraction framework to construct discriminative entity pair representations by integrating global semantic and local structural features. From the fused output of the joint ELMo-BERT embeddings, the hidden state corresponding to the special token [CLS] is used as the holistic representation of the entity pair. Global representation extraction is shown in Equation 9:

$$V_{pair} = h_{|CLS|} \tag{9}$$

This representation captures the global semantic association features of the entity pair. By retaining the [CLS] token's capability for modeling global relationships, it mitigates the bias where standalone ELMo tends to focus on local syntax while BERT emphasizes long-range dependencies. For an entity pair(e_1, e_2), the average hidden states of their respective textual segments are computed separately. Local Representation Extraction is shown in Equation 10 and 11:

$$V_{e_1} = \frac{1}{m_1} \sum_{i \in T_{e_1}} h_i \tag{10}$$

$$V_{e_2} = \frac{1}{m_2} \sum_{i \in T_{e_2}} h_i \tag{11}$$

where $T_(e_1)$ andT_{e_2} denote the token index sets for entities e_1 and e_2 , respectively, while m_1 and m_2 represent the corresponding token counts. The final representation of the entity pair is obtained by concatenating these two local representations, as shown in Equation 12:

$$V_{pair} = [V_{e_1}; V_{e_2}] \tag{12}$$

3.5.2 Similarity Calculation

(1) Design a multi-granularity similarity evaluation function for similarity computation, as shown in the Equation 13:

$$s(e_1, e_2) = \frac{V_{pair_1} \cdot V_{pair_2}}{||V_{pair_1}|| \cdot ||V_{pair_2}||} \tag{13}$$

Let V_{pair_1} and V_{pair_2} represent the entity pair vectors, with values ranging from $[-1, 1]$.

3.5.3 Unified Loss Function

To jointly optimize the embedding representations and the alignment decision, we design a unified loss function. This function consists of two parts: a contrastive learning loss L_align to pull together the representations of aligned entities and push apart those of non-aligned ones, and a regularization term L_threshold to guide the threshold learning.

Specifically, for a set of entity pairs (e_1, e_2, y), where $y = 1$ denotes alignment and y=0 denotes non-alignment, the loss function is defined as the Equation 14 and 15:

$$L = L_{align} + \lambda L_{threshold} \tag{14}$$

$$L_{align} = -\frac{1}{N} \sum_{(e_1, e_2, y)} [y \cdot loga(p(e_1, e_2)) + (1 - y) \cdot loga(1 - p(e_1, e_2))] \tag{15}$$

Here, $p(e_1, e_2) = \tilde{A}(s(e_1, e_2)/T)$ is the temperature-scaled alignment probability, see the Equation 16, and T is a learnable temperature parameter.

$$L_{threshold} = (\theta - s_{margin})^2 \tag{16}$$

Here, θ is a learnable threshold parameter, and s_{margin} is a preset margin value. This regularization term encourages the threshold to reside in a reasonable interval of the similarity distribution. λ is a coefficient balancing the two losses.

3.5.4 Entity Alignment Decision

In our framework, the alignment threshold θ is defined as a learnable model parameter, optimized jointly with the embedding layers via gradient descent. During inference, we directly use the converged value of θ as the decision threshold. The alignment decision rule is shown in the Equation 17:

$$align(e_1, e_2) = \begin{cases} 1, & \text{if } s(e_1, e_2) \geq \theta \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

This approach transforms threshold determination from a computationally expensive post-hoc grid search into an intrinsic part of the model's forward propagation, enabling end-to-end training and ensuring the threshold is co-optimized with the embeddings in the same semantic space.

3.6 Training Algorithm

The complete end-to-end training procedure of the proposed MulF-ELMo-BERT framework is summarized in Algorithm 1.

Algorithm 1 presents the complete end-to-end training procedure of the proposed MulF-ELMo-BERT framework. First, all model parameters are initialized, including ELMo, Chinese-BERT, multi-head attention, gating layer, and a learnable decision threshold θ. During each training epoch, for every batch in the training set D, the model performs a forward pass: it extracts multi-features for the entity pair (e_1, e_2), obtains embeddings via ELMo and Chinese-BERT, fuses these two types of embeddings through multi-head attention and a gating layer to produce the joint representation V_{pair}, and then computes the similarity score $s(e_1, e_2)$ along with the alignment probability $p(e_1, e_2)$.

Algorithm 1 : MulF-ELMo-BERT Training Procedure

Require: Training set D, learning rate η, margin s_{margin}, balance coefficient λ
Ensure: Trained model parameters (including threshold θ)

Initialize model parameters (ELMo, BERT, MHA, Gating Layer, threshold θ)
for epoch = 1 to N **do**
  for each batch in D **do**
    Forward Pass
    Extract multi-features for entity pair (e_1, e_2)
    Obtain embeddings via ELMo and Chinese-BERT
    Fuse embeddings via Multi-Head Attention and Gating Layer
    Compute entity pair representation V_{pair}
    Compute similarity score $s(e_1, e_2)$
    Compute alignment probability $p(e_1, e_2)$
    Loss Computation
    Compute alignment loss L_{align}
    Compute threshold regularization loss $L_{threshold}$
    Compute total loss $L = L_{align} + \lambda L_{threshold}$
    Backward Pass
    Compute gradients ∇L
    Update all parameters (including θ) using optimizer
  end for
end for

Subsequently, the alignment loss L_{align} and threshold regularization loss $L_{threshold}$ are calculated, and the total loss L is obtained as a weighted sum controlled by coefficient λ. Finally, gradients are computed via backpropagation, and all parameters (including threshold θ) are updated using an optimizer.

4 Experiments

4.1 Datasets

To verify the validity of the proposed alignment method, the dataset required for the experiment is constructed. First, some entities of the cultural relics' entity corpus have been directly extracted from the Shaanxi History Museum and the National Cultural Relics Collection Directory.

Due to the heavy workload and low accuracy of manual annotation data, this paper extracts entities from the Baidu Encyclopedia and Chinese Wikipedia to automatically generate training data sets by using heuristic rules with the help of some structured information in the encyclopedia.

The heuristic rules for positive example selection in the training dataset are as follows:

(1) Two article entries have the same and unique names, that is, the two article entries have no synonyms or aliases.

(2) Two articles with the same title, and the content similarity of the article must exceed 95%.

(3) The category labels of the two articles are exactly the same, and the content similarity of the two articles is over 95%.

(4) Two articles have the same title and the same category tag, they will be listed as referring to the same entity.

In addition, there is the function of synonym digestion in the encyclopedia. That is, an article name has an alias, but will be automatically replaced with the standard entity name when looked up on the encyclopedia site, e.g., "shan Li Bo" is the alias of "shaanxi history museum". When "shan Li Bo" is received as input, it will be replaced by the "Shaanxi History Museum" automatically, and the information about the Shaanxi Provincial History Museum is returned.

Many rules are used for negative example selection in the training dataset. Except for rules that generate positive examples, all other rules can generate negative examples. Enumerate the main

Table 1: The dataset of cultural relic entities

Entity Types	Baidu Encyclopedia entity number	Wikipedia entity number
CRN	1876	1700
UL	1785	1857
HD	1562	1800
MC	1500	1700

heuristic rules used for negative example generation. (1) The titles of two articles are different, and the content similarity of the two articles is less than 50%. (2) The category labels of the two articles are completely different, and the content similarity of the two articles is less than 50%. (3) The two articles are in different fields.

The descriptions of four types of cultural relic entities, cultural relic' s name (CRN), unearthed location (UL), historical dynasty (HD), museum collection (MC), in the experimental dataset are shown in Table 1.

We randomly split the dataset into training (70%), validation (15%), and test (15%) sets. To ensure robustness, we perform 5-fold cross-validation and report the average performance. The same split is used for all baseline comparisons.

4.2 Experiment settings

4.2.1 Evaluation Metrics

To quantify the alignment accuracy and ranking quality of the model in Entity Alignment (EA) tasks, this paper employs two widely recognized evaluation metrics: Hits@N and MRR (Mean Reciprocal Rank). Hits@N directly reflects the model's alignment capability under different cutoff thresholds, making it suitable for evaluating strict matching (Hits@1) and fuzzy matching (Hits@10) scenarios in entity alignment tasks. MRR mitigates the impact of extreme rankings (such as extremely low rankings) through a reciprocal ranking mechanism, providing a more robust measure of the model's global ranking performance, especially in scenarios with a large number of candidate entities. Their definitions and calculation methods are as follows.

(1)Hits@N

Hits@N is used to measure the probability that the model correctly predicts the aligned entity within the top N positions of the ranked list of candidate entities. For each entity pair (e_1, e_2) in the test set, if the position t of its correctly aligned entity e_j in the ranked list satisfies $t \leq N$, it is counted as a valid prediction $\delta(t \leq N) = 1$; otherwise, it is considered invalid $\delta(t > N) = 0$. The mathematical expression for Hits@N is shown in Equation 18:

$$Hits@N = \frac{1}{|T|} \sum_{(e_1,e_2)\in T} \delta(t \leq N), N \in \{1, 10\} \tag{18}$$

where T represents the set of aligned entity pairs in the test set, and $|T|$ is the size of the set. Hits@1 reflects the model's accuracy in top-1 predictions (strict alignment capability), while Hits@N evaluates the model's recall capability under relaxed conditions (fault tolerance and robustness). A higher metric value indicates better model performance.

(2) MRR (Mean Reciprocal Rank)

MRR assesses the model's ranking quality by calculating the average of the reciprocal ranks (RR) of the correctly aligned entities in the ranked lists. For each test sample(e_1, e_2), if the rank of its correct entity e_j is t, its RR value is $\frac{1}{t}$; if the correct entity does not appear in the top n positions, the RR is recorded as 0. The mathematical expression for MRR is shown in Equation 19:

$$MRR = \frac{1}{|T|} \sum_{(e_1,e_2)\in T} \frac{1}{t} \tag{19}$$

The range of MRR values is (0,1], where 1 indicates that all correct entities are ranked first. A higher value signifies that the model ranks the correct entities more prominently, indicating better

overall performance.

4.2.2 Baselines

To prove the effectiveness of MulF-ELMo-BERT proposed in this paper, the experimental results are compared with the following baselines.

- Entity Alignment Methods for Encyclopedic Knowledge Bases:
  - Gerlach et al. [3] proposed a human-computer collaborative entity linking system.
  - Gabrilovich et al. [4] introduced the Explicit Semantic Analysis (ESA) method.
  - Lin et al. [5] proposed an end-to-end framework for entity recognition and disambiguation.
  - Deng et al. [19] proposed an event entity alignment method based on event elements.
- Embedding-Based Entity Alignment Methods:
  - TransE [20]. TransE aims to make the embedding of the tail entity t in a relational triple as close as possible to the sum of the embeddings of the head entity h and the relation r.
  - GloVe [21]. Yoon et al. investigated the use of semantic features from GloVe and lexical features from Bigram to evaluate the accuracy of dual alignment.
  - Word2vec [22]. Kolyvakis et al. studied the use of word2vec word vector sets to verify the impact of initial pre-trained word vectors on model performance.
  - ELMo [23]. Shahbazi et al. proposed a new local entity disambiguation system.
- Deep Learning-Based Entity Alignment Methods:
  - GCN [9]. Tam et al. proposed an end-to-end unsupervised entity alignment framework.
  - RDGCN [24]. Wu et al. proposed a relation-aware dual graph convolutional network.
  - RNM [25]. Zhu et al. proposed a relation-aware neighborhood matching model.
  - BERT [26]. Bai et al. proposed a new entity alignment method based on BERT.
  - RoBERTa-zh [27]. An optimized BERT pre-training approach proposed by Liu et al., which demonstrates strong performance on Chinese tasks.
  - ERNIE [28]. A knowledge-enhanced pre-trained model proposed by Baidu, which improves semantic representation by incorporating knowledge graphs.
  - XLNet-Chinese [29]. Employing a generalized autoregressive pre-training approach that overcomes the limitations of the masked independence assumption in BERT.

All models are tuned on the validation set via grid search. Key hyperparameters include learning rate 1e-5, 2e-5, 5e-5, batch size 16, 32, and ELMo window size 3, 5, full. We fix the random seed to 42 for reproducibility.

4.3 Experimental results and analysis

4.3.1 Comparison experiment of entity alignment for cultural relics

To verify the overall performance of the MulF-ELMo-BERT model proposed in this paper in the task of entity alignment for cultural relics, the comparative experiments between the proposed model and baselines are executed. In this experiment, four types of cultural relic entities (CRN, UL, HD, MC) are selected for the comparative experiments. The experimental results are shown in Table 2.

As shown in Table 2, in terms of the Hits@1 metric, the MulF-ELMo-BERT model demonstrates significant improvements over the strongest baseline models. Compared with the second-best performer ERNIE, our model achieves improvements of 2.8%, 2.8%, 3.4%, and 3.1% across the four entity types (CRN, UL, HD, MC), respectively. To evaluate the statistical significance of these performance improvements, we conducted paired t-tests between MulF-ELMo-BERT and the best baseline

Table 2: Experimental results of entity alignment for cultural relics

Entity	CRN			UL			HD			MC		
Models	Hits @1	Hits @10	MRR	Hits @1	Hits @10	MRR	Hits @1	Hits @10	MRR	Hits @1	Hits @10	MRR
Martin et al. [3]	42.2	60.2	0.523	49.3	62.6	0.565	45.7	67.8	0.551	47.4	63.5	0.586
Gabri et al. [4]	48.6	63.7	0.557	48.7	65.7	0.592	47.5	69.2	0.587	48.5	64.7	0.597
Lin et al. [5]	53.1	69.4	0.596	56.5	73.1	0.633	53.1	71.9	0.602	53.1	69.4	0.621
Deng et al. [19]	49.7	68.6	0.572	52.1	69.5	0.617	51.3	68.7	0.598	52.6	68.8	0.617
Trans E [20]	52.3	70.1	0.678	55.6	70.9	0.694	54.6	71.1	0.685	55.5	71.5	0.695
GloVe [21]	46.8	65.8	0.645	49.8	69.0	0.665	47.5	68.6	0.667	48.8	69.7	0.658
word2vec [22]	52.5	68.5	0.662	54.2	70.3	0.681	53.6	69.0	0.688	53.4	70.7	0.693
ELMo [23]	59.9	74.4	0.703	62.7	76.8	0.725	61.7	79.3	0.719	62.2	75.1	0.734
GCN [9]	65.4	80.3	0.721	67.4	82.6	0.766	66.4	85.7	0.742	67.6	81.3	0.742
RDGCN [24]	66.2	83.9	0.756	69.9	85.9	0.792	67.8	86.1	0.774	68.4	84.5	0.779
RNM [25]	68.7	85.6	0.773	71.5	87.4	0.817	69.5	91.3	0.801	70.7	86.2	0.782
Bert [26]	71.6	87.8	0.792	74.8	89.6	0.838	72.2	93.4	0.820	73.5	88.3	0.835
RoBERTa-zh [27]	72.8	88.5	0.805	75.6	90.3	0.851	73.5	94.1	0.835	74.2	89.2	0.848
ERNIE [28]	73.5	89.2	0.818	76.3	91.1	0.862	74.1	94.8	0.842	75.1	90.3	0.856
XLNet-Chinese [29]	72.1	88.1	0.811	75.1	90.0	0.855	73.2	93.9	0.838	74.0	89.0	0.851
MulF-ELMo-BERT	76.3	92.5	0.845	79.1	93.2	0.883	77.5	96.8	0.872	78.2	93.7	0.894

Note: All performance metrics in this table are the mean values after 5 independent runs with random seeds. The standard deviation ranges are as follows: Hits@1: 0.3%–0.6%, Hits@10: 0.2%–0.4%, MRR: 0.003–0.007. The improvements of MulF-ELMo-BERT over the best baseline (BERT) are statistically significant for all entity types (paired t-test).

models (ERNIE, RoBERTa-zh, and XLNet-Chinese). The results indicate that the improvements are statistically significant (p-value < 0.05) across all four entity types. Notably, it achieves 77.5% on the HD type, showcasing superior precision in entity semantic matching.

The baseline comparison experiment indicates that the end-to-end joint training framework significantly enhances model performance across all entity types. While advanced pre-trained models like ERNIE, RoBERTa-zh, and XLNet-Chinese show competitive results, our approach of integrating ELMo's contextual awareness with Chinese-BERT's deep semantic representations through the proposed gating mechanism and jointly optimizing the alignment threshold proves to be more effective for cultural relic entity alignment. MulF-ELMo-BERT achieves more precise semantic matching and robust alignment performance, providing an efficient and precise entity alignment tool for the digital preservation of cultural heritage.

To evaluate the statistical significance of the performance improvement, we conducted paired t-tests between MulF-ELMo-BERT and the top baseline models (ERNIE, RoBERTa-zh, and XLNet-Chinese) on the test set. The results indicate that the improvements in the Hits@1 metric are statistically significant (p-value < 0.05) across all four entity types and against all three strong baselines. This demonstrates that the performance gain from our model is not due to chance and validates the effectiveness of our proposed fusion framework.

4.3.2 Experiment on the Advantages of Fused Word Embeddings

In the task of cultural relic entity alignment, to explore the performance of different embedding models and their fusion strategies, we designed and conducted an experiment on the advantages of fused word embeddings. This experiment selected three single-model word embeddings: ELMo embedding, BERT-base embedding, and Chinese-BERT embedding, as well as different combinations such as ELMo + BERT-base fused embedding and ELMo + Chinese-BERT fused embedding. The experimental results are shown in Table 3.

As shown in Table 3, the proposed MulF-ELMo-BERT framework demonstrates enhanced performance across various scenarios of cultural relic entity alignment under the end-to-end training paradigm. In the CRN task, it achieves more precise differentiation with Hits@1 (76.3%) and MRR (0.845), significantly surpassing the best-performing single model, Chinese-BERT. For UL entities, the framework better handles geographical hierarchies and reduces place name confusion. In HD align-

Table 3: Experiment on the Advantages of Fused Word Embeddings

Entity	CRN			UL			HD			MC		
Word embedding	Hits @1	Hits @10	MRR	Hits @1	Hits @10	MRR	Hits @1	Hits @10	MRR	Hits @1	Hits @10	MRR
ELMo	51.9	78.4	0.693	56.1	80.8	0.735	55.9	84.3	0.714	55.8	79.9	0.734
BERT-base	63.1	85.8	0.792	68.0	83.6	0.838	66.4	89.4	0.820	66.5	84.3	0.835
Chinese-BERT	63.8	87.3	0.802	68.4	87.7	0.844	67.2	90.5	0.831	67.6	87.4	0.841
ELMo+BERT-base	70.7	89.8	0.816	70.6	89.3	0.862	69.5	91.7	0.843	69.4	89.5	0.868
MulF-ELMo-BERT	76.3	92.5	0.845	79.1	93.2	0.883	77.5	96.8	0.872	78.2	93.7	0.894

Note: All performance metrics in this table are the mean values after 5 independent runs with random seeds. The standard deviation ranges are consistent with those reported in Table 2.

ment, it shows improved capability in distinguishing temporal ranges and character variants. For MC tasks, the enhanced fusion strategy enables more accurate ranking of collection institutions.

The limitations of single-model embeddings remain evident in the improved framework: ELMo still lacks deep contextual semantic modeling, BERT-base continues to suffer from Chinese optimization issues, and Chinese-BERT, while performing best among single models, still cannot effectively capture fine-grained character-level features.

The superiority of the fusion strategy is further validated: Compared with ELMo+BERT-base, the enhanced MulF-ELMo-BERT achieves an average increase of 4.8% in Hits@1 and 2.7% in Hits@10 across all entity types. The gating mechanism demonstrates stronger feature complementarity under joint optimization, effectively balancing the dynamic contextual adjustments of ELMo with the global semantic representations of Chinese-BERT, while maintaining parameter efficiency. This substantiates the effectiveness of synergistic embedding fusion within the end-to-end learning framework.

4.3.3 Influence of entity features on the entity alignment performance for cultural relics

In the research on cultural heritage digitization, entity alignment serves as a core task for integrating multi-source heterogeneous data and constructing high-quality knowledge graphs. However, the alignment performance of cultural relic entities is often constrained by expression ambiguities, semantic dependencies, or data sparsity. To explore the impact of different attribute combinations on entity alignment, this study conducted comparative experiments on four types of entities based on seven fusion strategies incorporating structured attributes (A), text summaries (S), and full-text contexts (C). The experimental results are presented in Figure 2.

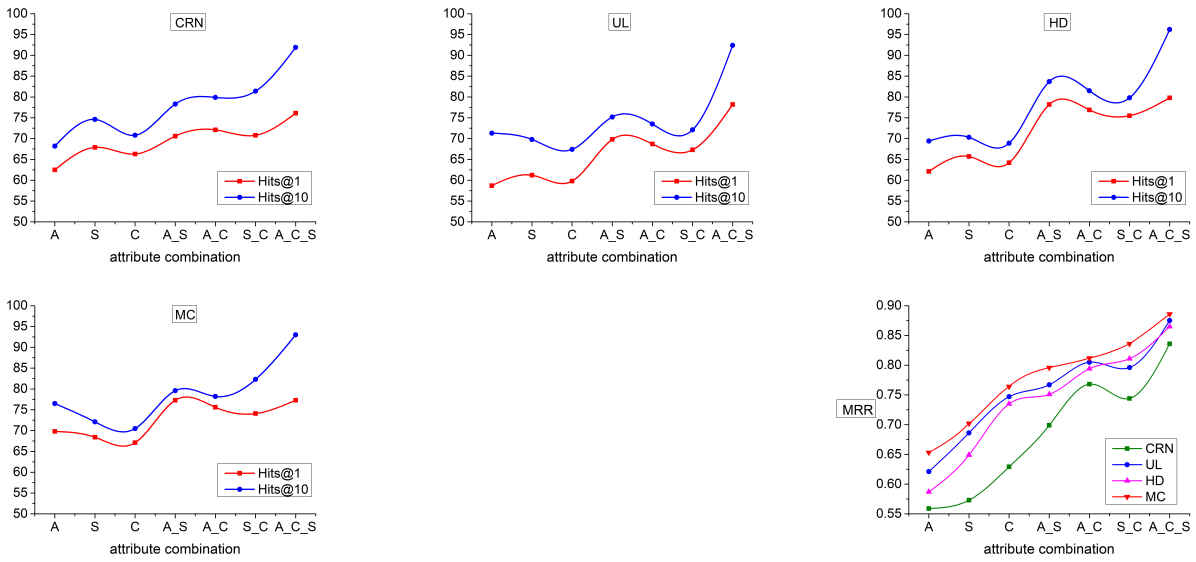


Figure 2: The influence of entity features on cultural relic entity alignment

In the Figure 2, A, S, C, A_S, A_C, S_C, and A_C_S represent attributes (A), text summaries (S), full-text contexts (C), the fusion of attributes (A) and text summaries (S), the fusion of attributes (A) and full-text contexts (C), the fusion of text summaries (S) and full-text contexts (C), and the fusion of attributes (A), text summaries (S), and full-text contexts (C), respectively.

As shown in Figure 2, in the cultural heritage entity alignment experiments under the end-to-end training framework, there are significant performance differences among various attribute combinations.

Regarding single attributes, the A performs well in the HD and MC categories. This is because these types of entities rely on precise matching of structured attributes. However, it underperforms in the CRN and UL categories, as the presence of aliases or descriptive discrepancies in names leads to insufficient attribute coverage. The S outperforms the A in the CRN category since the summaries contain synonyms for names or contextual information. In contrast, it performs poorly in the UL and HD categories because geographical and dynastic information is simplified or omitted. The C generally exhibits lower performance than the S with Hits@1 for all entity types not exceeding 72%, due to interference from noise in the full text.

Among dual-attribute fusions, the combination of "A_S" shows significant improvement in the CRN category; "A_C" demonstrates notable enhancement in the UL category; while "S_C" yields limited improvement. The tri-attribute fusion "A_S_C" delivers the best performance, achieving Hits@1 values of 76.1%, 78.2%, and 79.8% for CRN, HD, and UL, respectively, as it can integrate the advantages of multiple attributes.

In terms of entity types, HD and MC are sensitive to structured attributes, with dual-attribute or tri-attribute fusions resulting in improvements exceeding 5%. CRN and UL rely on textual descriptions and context, with tri-attribute fusions achieving improvements exceeding 10%. This fully demonstrates the effectiveness of multi-attribute fusion in the task of cultural relic entity alignment.

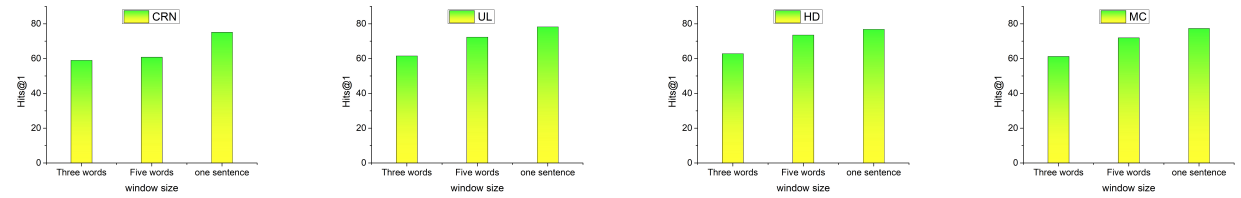
4.3.4 The impact of ELMo's window size on cultural relic entity alignment

ELMo, as a pre-trained language model capable of capturing contextual semantic features, has its ability to capture local contextual information in text directly influenced by the setting of its window size. To explore the impact of ELMo's window size on the task of cultural relic entity alignment, we conducted a series of comparative experiments. We tested the model's performance when ELMo's window size was set to 3 words, 5 words, and the entire sentence, respectively, as shown in Figure 3.

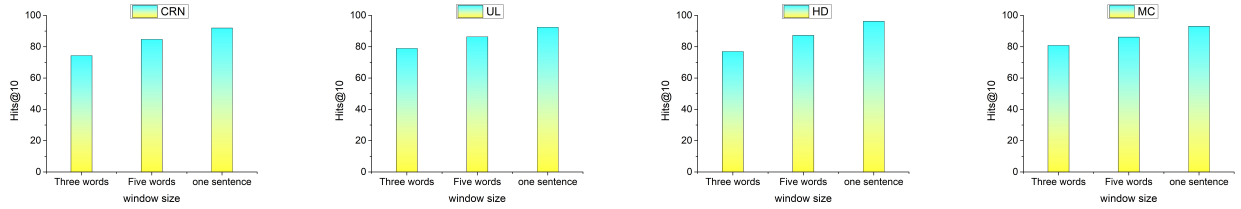
As shown in Figure 3, in the task of aligning cultural relics entities using the end-to-end trained model, the size of the ELMo window maintains a significant impact on the model's performance. As the window size increases from 3 words to the entire sentence, the Hits@1, Hits@10, and MRR indicators for each type of cultural relic entity continue to show a clear upward trend, with performance gains being more pronounced in the enhanced framework. Among them, the model performs best under the full sentence window, achieving the highest scores of 76.3% (CRN), 79.1% (UL), 77.5% (HD), and 78.2% (MC) in Hits@1 across all four entity types, indicating that expanding the window can better enhance the model's ability to capture contextual semantics and improve alignment accuracy within the joint optimization framework. The 5-word window performs second best, achieving approximately 1.5-2.0% lower Hits@1 scores compared to the full sentence configuration, while the 3-word window shows the most limited performance with Hits@1 scores remaining below 74% for all entity types. This consistent pattern suggests that in the alignment of cultural relic entities, due to the special composition and semantic complexity of cultural relic entity names, fully considering long-distance contextual information remains crucial for achieving optimal performance in the improved model architecture.

4.3.5 Comparison of task metrics under different numbers of epochs

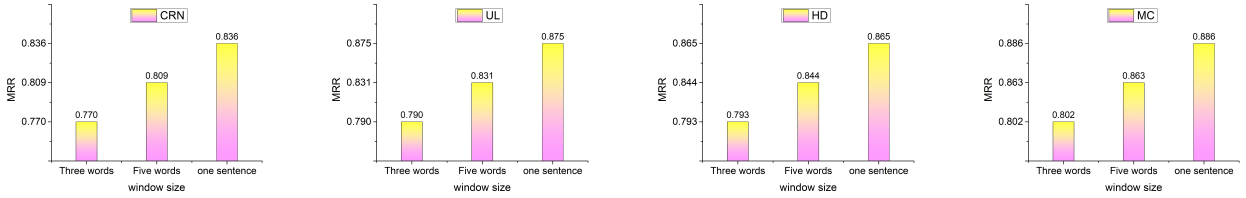
Due to semantic and structural differences, tasks involving different types of entities exhibit varying sensitivities to the number of training epochs. To investigate the impact of the number of epochs on various cultural relic entity alignment tasks, we selected four typical tasks and compared their



(a) The impact of window size on Hits@1



(b) The impact of window size on Hits@10



(c) The impact of window size on MRR

Figure 3: The impact of ELMo's window size on cultural relic entity alignment

performance metrics across different epoch counts (1, 5, 10, 15, and 20). The experimental results are shown in Figure 4.

As shown in Figure 4, when there was only 1 epoch, all task metrics dropped significantly, with Hits@1 for both CRN and UL decreasing by approximately 7%. This indicates that before 10 epochs, the model had not fully learned the features of cultural relic entities and had insufficient learning ability for tasks dependent on long texts. At 15 epochs, the HD task showed slight improvements in Hits@1 and MRR because historical dynasty names have strong semantic associations. The metrics for CRN, UL, and MC tasks fluctuated slightly (around ±0.5%), indicating that the model was approaching convergence at 10 epochs and that excessive training yielded limited benefits. At 20 epochs, with the adoption of a dynamic learning rate, the metrics were close to those at 15 epochs, suggesting that this strategy could alleviate the vanishing gradient problem but that the cultural relic entity alignment task was less sensitive to long-term training.

The HD task was the most sensitive to the number of epochs, with the highest gains observed at 15 epochs, as optimizing the semantic associations of dynasty names required contextual information. The MC task was the most robust, with metrics remaining stable in the later stages, as its features were simple and easy to converge. The CRN and UL tasks exhibited intermediate sensitivity, necessitating a balance between local and global feature learning.

4.3.6 Comparison of task metrics under different learning rate strategies

Different entity tasks exhibit varying sensitivities to learning rates due to differences in semantics and data structures. To investigate the influence mechanism of learning rate strategies, we selected four typical entity tasks and employed four learning rate strategies for training and evaluation. The experimental results are presented in Table 4.

As shown in Table 4, the learning rate significantly affects model convergence. With a high learning rate (5e-5), the proposed end-to-end trained model achieves the best performance across all entity types, with the HD task reaching 77.5% Hits@1 and 0.872 MRR, demonstrating the framework's capability to effectively capture complex semantic associations. The CRN task benefits substantially,

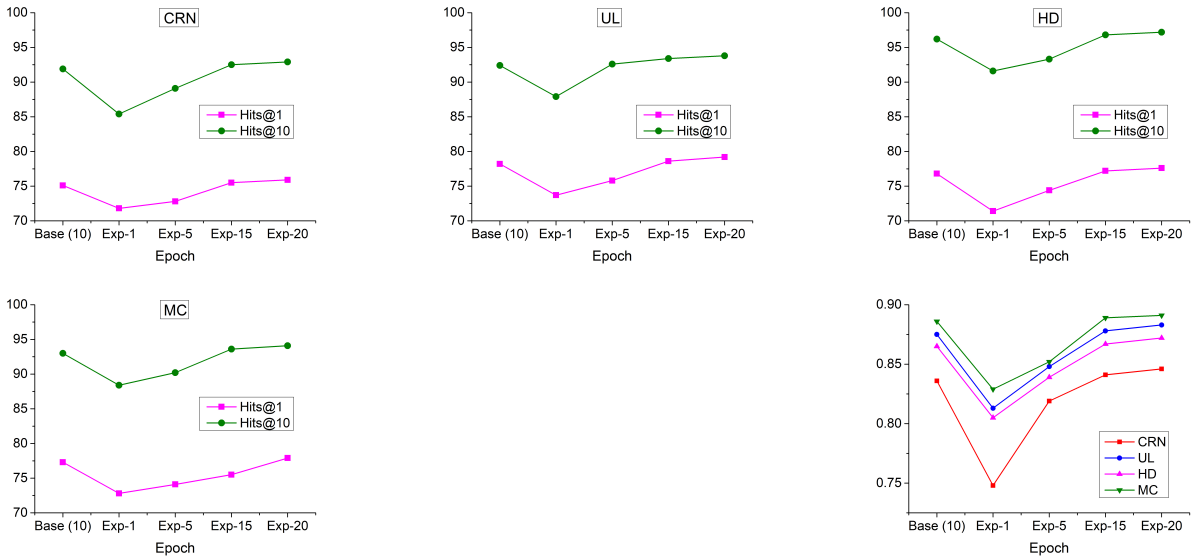


Figure 4: Comparison of task metrics under different numbers of epochs

Table 4: Comparison of task metrics under different learning rate strategies

Entity	CRN			UL			HD			MC		
Learning Rate	Hits @1	Hits @10	MRR	Hits @1	Hits @10	MRR	Hits @1	Hits @10	MRR	Hits @1	Hits @10	MRR
Base(2e-5)	74.8	91.5	0.761	77.7	91.9	0.800	76.3	95.7	0.790	76.8	93.4	0.867
High-LR(5e-5)	76.3	92.5	0.845	79.1	93.2	0.883	77.5	96.8	0.872	78.2	93.7	0.894
Low-LR (1e-5)	74.1	91.1	0.740	77.4	91.6	0.779	75.9	95.3	0.769	76.5	92.5	0.857
Linear-Decay	75.2	91.8	0.799	78.1	92.4	0.843	76.7	96.1	0.831	77.2	93.8	0.872

Note: All performance metrics in this table are the mean values after 5 independent runs with random seeds. The standard deviation ranges are consistent with those reported in Table 2.

achieving 76.3% Hits@1, while UL reaches 79.1% Hits@1, indicating robust handling of textual descriptions. Even the relatively simpler MC task shows improved performance at 78.2% Hits@1. A low learning rate (1e-5) causes all task metrics to decline, with the HD task being most affected because the slow parameter updates make it difficult to learn complex semantics. In terms of task characteristics, the HD task is most sensitive to the learning rate, benefiting greatly from a high learning rate as it requires contextual semantic understanding. The MC task is more robust, with simple museum name features and minimal complex semantic interactions. The CRN and UL tasks have moderate sensitivity, and the proposed framework consistently delivers strong performance across different learning rate strategies.

4.3.7 Analysis of Ablation Study Results

This section systematically validates the necessity of the proposed core components through ablation studies. The analysis of ablation study results is shown in Table 5.

The ablation studies in Table 5 systematically evaluate our core contributions, where "w/o Joint Threshold" removes the learnable threshold module, "w/o Gating Fusion" eliminates the adaptive fusion mechanism, and "w/o Both" ablates both key components.

The results confirm the effectiveness and synergistic effects of our proposed mechanisms. Removing the joint threshold learning (reverting to grid search) caused an average 1.0% decrease in Hits@1, demonstrating that learnable parameters optimize decision boundaries in semantic space. More importantly, eliminating the gating mechanism led to a 0.9% performance drop, which can be directly attributed to the loss of adaptive feature weighting.

Further analysis reveals that in the complete model, the gating values G_i dynamically adjust based

Table 5: Ablation Study Results

Entity	CRN			UL			HD			MC		
Model Variant	Hits @1	Hits @10	MRR	Hits @1	Hits @10	MRR	Hits @1	Hits @10	MRR	Hits @1	Hits @10	MRR
MulF-ELMo-BERT	76.3	92.5	0.845	79.1	93.2	0.883	77.5	96.8	0.872	78.2	93.7	0.894
w/o Joint Threshold	75.1	91.9	0.836	78.2	92.4	0.875	76.8	96.2	0.865	77.3	93.0	0.886
w/o Gating Fusion	74.8	92.1	0.838	78.5	92.8	0.879	77.0	96.5	0.868	77.8	93.4	0.890
w/o Both	73.9	91.5	0.829	77.6	92.1	0.871	76.2	95.9	0.861	76.9	92.8	0.884

Note: All performance metrics in this table are the mean values after 5 independent runs with random seeds. The standard deviation ranges are consistent with those reported in Table 2.

on input characteristics—they are generally higher (favoring ELMo) when processing rare characters in cultural relic names, and lower (favoring BERT) when handling structured information like historical dynasties. This intelligent, input-dependent fusion capability represents a significant advantage unattainable by any single model or fixed fusion strategy. Notably, the simultaneous removal of both components resulted in a 1.8% performance degradation, exceeding the sum of individual removals and highlighting the deep coupling between adaptive threshold learning and feature fusion, thereby validating the overall superiority of our end-to-end joint learning framework.

4.4 Efficiency Analysis

To comprehensively evaluate the practicality of the proposed model, this section conducts an efficiency analysis of the MulF-ELMo-BERT framework from both theoretical complexity and practical runtime perspectives.

Theoretical Complexity Analysis: The computational cost of our framework primarily stems from the forward passes of the ELMo and BERT encoders. Given a sequence length n, the complexity of ELMo (based on BiLSTM) is $O(L_e \cdot n^2)$, where L_e is the number of LSTM layers. The complexity of BERT (based on Transformer) is similarly $O(L_b \cdot n^2)$, where L_b is the number of Transformer layers. The fusion layer (Multi-Head Attention and Gating) has a complexity of $O(n^2 \cdot d \cdot h)$, where d is the embedding dimension and h is the number of attention heads. Since d and h are fixed constants, the overall training complexity is dominated by the square of the sequence length. However, during the inference stage, embedding vectors for all entities can be precomputed and indexed offline. Consequently, the online alignment decision only requires computing the cosine similarity between two vectors, which has a constant complexity $O(1)$, ensuring the framework's applicability to large-scale knowledge bases.

Practical Runtime Comparison: We report the average inference time (milliseconds per entity pair) on the test set and the parameter count for various baseline models and our model in Table 6. All experiments were conducted under the same hardware environment (NVIDIA Tesla V100 GPU) with the same batch size.

Table 6: Model Efficiency Comparison

Model	Params (M)	Inference Time (ms/pair)
TransE	5.2	1.2 ± 0.1
GCN	8.7	3.5 ± 0.3
BERT	110.2	15.8 ± 1.2
MulF-ELMo-BERT	125.6	18.3 ± 1.5

The results in Table 6 indicate that our proposed model has a slightly higher parameter count and inference time than the pure BERT model, which represents the inherent overhead introduced by the dual-encoder fusion architecture. In terms of training efficiency, the average time for MulF-ELMo-BERT to complete one training epoch on the full dataset is 3.2 hours, which remains on the same order of magnitude as the 2.4 hours required by the BERT baseline. While these computational costs are moderately increased, the overhead is within an acceptable range, and the significant performance gains demonstrated in Section 4.3 convincingly establish a favorable trade-off between effectiveness

and practicality in our framework.

5 Conclusion

To address the challenge of heterogeneous names but isomorphic semantics in entity alignment across multi-source knowledge bases caused by manual input variations, this paper proposes a general joint-embedding framework named MulF-ELMo-BERT. We present an end-to-end joint learning framework and demonstrate the effectiveness of its core components through ablation studies. The framework effectively tackles entity alignment challenges in domains with dense specialized terminology through multi-dimensional feature fusion and adaptive semantic calibration. Its effectiveness has been validated specifically in the cultural heritage domain.

The framework integrates multi-dimensional features such as entity names, attributes, summaries, and full-text content, enabling comprehensive extraction of entity features at four levels: characters, words, sentences, and paragraphs. It effectively filters out weakly relevant entities and breaks through the limitations of semantic representation that rely solely on a single feature. Considering the dense presence of rare characters and specialized terms in cultural relic entity names, the framework incorporates the ELMo context-aware word embedding model, which can dynamically adjust the vector representation of rare words based on their context, significantly enhancing semantic adaptability. Meanwhile, a Chinese BERT model with an integrated whole-word masking strategy is adopted to avoid interference from local co-occurrence and strengthen the ability to capture term associations. More importantly, the framework deeply fuses ELMo embeddings with Chinese BERT model embeddings as entity embeddings, effectively compensating for the limitations of single embedding models in complex semantic representation. Additionally, it improves matching accuracy through adaptive calibration of cosine similarity using a dynamic threshold determination method, enabling precise alignment of cultural relic entities.

Experimental results demonstrate that the proposed method can efficiently capture multi-level semantic features of entities and exhibits excellent performance in cultural relic entity alignment tasks, providing a feasible solution for entity alignment in cultural relic knowledge fusion.

We indicate that while the method is primarily designed for the cultural relic domain, its technical framework (e.g., multi-feature fusion, adaptable pre-trained models) demonstrates domain generality. It can be extended to fields like medicine or geography by employing domain-specific pre-trained models (e.g., BioBERT, GeoBERT) and adjusting feature weights. However, this study has several limitations: (1) The method's effectiveness in the cultural relic domain relies on the quality of domain texts, and its performance may be limited in domains with scarce annotated data. (2) The current approach mainly handles Chinese text, and its applicability to cross-lingual entity alignment requires further validation.

Our future work will focus on the following directions: First, we will explore cross-domain and cross-lingual generalization. Specifically, we plan to apply our framework to domains with rich terminologies, such as medicine and geography, by leveraging domain-specific pre-trained models (e.g., BioBERT, SciBERT) and investigating self-adaptive mechanisms for feature weighting across domains to validate and enhance the method's universality. Concurrently, we will construct multilingual cultural relic knowledge datasets and investigate joint embedding strategies that incorporate multilingual pre-trained models (e.g., XLM-RoBERTa) to address the challenges of cross-lingual entity alignment. Secondly, we will commit to the refinement and optimization of the technical approach. To tackle the challenge in domains with scarce data, future research will focus on few-shot learning and self-supervised learning strategies to minimize reliance on large-scale annotated data. Furthermore, we intend to integrate more profound semantic relationship modeling techniques, such as Graph Neural Networks, to capture the complex associations between entities within knowledge graphs, thereby further improving alignment accuracy and robustness.

Funding

This research was funded by the Ministry of Education in China Project of Humanities and Social Sciences (Project Title: Research on Intelligent Question Answering and Knowledge Recommenda-

tion for Smart Museums Based on Cultural Relics Knowledge Graph, grant number: 23YJA870016）; the Shaanxi Provincial Philosophy and Social Sciences Research General Project (Project Title: Research on the Pathway of Museum Intelligent Navigation and Cultural Dissemination in Shaanxi Empowered by Cultural Heritage Knowledge Graphs, grant number: 2026YB0087)；The Science Research Program of Shaanxi Provincial Department of Education (Project Title: Research on Intelligent Recognition Algorithm for Traffic Sign Images for Autonomous Driving,grant number: 25JK0406）.

Author contributions

Literature research and Conceptualization, Min Zhang and Luya Yang; Data curation and data analysis, Yaxian Gao; Methodology, Min Zhang; Software, Luya Yang; Writing-original draft preparation, Min Zhang; Writing-review and editing, Min Zhang and Yong Ren.

Conflict of interest

The authors declare no conflict of interest.

References

[1] Zhang, R.; Trisedya, B.D.; Li, M.; Jiang, Y.; Qi, J. (2022). A benchmark and comprehensive survey on knowledge graph entity alignment via representation learning, *The VLDB Journal*, 31(5), 1143–1168, 2022.

[2] Hyvönen, E.; Rantala, H. (2019). Knowledge–based Relation Discovery in Cultural Heritage Knowledge Graphs, *DHN 2019: Proceedings of the Digital Humanities in the Nordic Countries 4th Conference, Copenhagen, Denmark, March 5–8, 2019*, 230–239, 2019.

[3] Gerlach, M.; Miller, M.; Ho, R.; Harlan, K.; Difallah, D. (2021, October). Multilingual entity linking system for wikipedia with a machine-in-the-loop approach, *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, 3818–3827, 2021.

[4] Gabrilovich, E.; Markovitch, S. (2007). Computing semantic relatedness using Wikipedia–based explicit semantic analysis, *IJCAI'07: Proceedings of International Joint Conference on Artificial Intelligence, Macao, August 10–16, 2007*, 1606–1611, 2007.

[5] Lin, L.; Hao, T. (2024). An end-to-end entity recognition and disambiguation framework for identifying Author Affiliation from literature publications, *Proceedings of the Fourth Workshop on Scholarly Document Processing (SDP 2024)*, 120–129, 2024.

[6] Zhang, F.; Li, J.; Cheng, J. (2023). Improving entity alignment via attribute and external knowledge filtering, *Applied Intelligence*, 53(6), 6671–6681, 2023.

[7] Yang, K.; Liu, S.; Zhao, J.; et al. (2020). COTSAE: co-training of structure and attribute embeddings for entity alignment, *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(03), 3025–3032, 2020.

[8] Wang, L.; Lu, L.; Jiang, N. (2011). A study of leaf modeling technology based on morphological features, *Mathematical and Computer Modelling*, 54(3-4), 1107–1114, 2011.

[9] Tam, N.T.; Trung, H.T.; Yin, H.; et al. (2020). Entity alignment for knowledge graphs with multi-order convolutional networks, *IEEE Transactions on Knowledge and Data Engineering*, 34(9), 4201–4214, 2020.

[10] Wang, L.; He, Y.; Zhou, S.; Lu, L. (2013). Modeling and visualization based on morphological features of leaf vein, *Sensor Letters*, 11(6-7), 1288–1292, 2013.

[11] Li, Y.; Xu, T.; Sun, Z.; et al. (2024). Entity Alignment Through Joint Utilization of Multiple Pretrained Models for Attribution Relationship, *2024 IEEE 11th International Conference on Cyber Security and Cloud Computing (CSCloud)*, IEEE, 1–6, 2024.

[12] Sun, Y.; Lee, Y. (2024). Embedding-based Two-Stage Entity Alignment for Cross-Lingual Knowledge Graphs, *Journal of Information Science and Engineering*, 40(2), 2024.

[13] Zhong, Z.; Zhang, M.; Fan, J.; et al. (2022). Semantics driven embedding learning for effective entity alignment, *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, IEEE, 2127–2140, 2022.

[14] Ge, C.; Liu, X.; Chen, L.; et al. (2021). Make it easy: An effective end-to-end entity alignment framework, *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, 777–786, 2021.

[15] Liu, Z.; Cao, Y.; Pan, L.; et al. (2020). Exploring and evaluating attributes, values, and structures for entity alignment, *arxiv preprint arxiv:2010.03249*, 2020.

[16] Paganelli, M.; Tiano, D.; Guerra, F. (2024). A multi-facet analysis of bert-based entity matching models, *The VLDB Journal*, 33(4), 1039–1064, 2024.

[17] Huang, J.; Li, T.; Jia, Z.; et al. (2016). Entity alignment of Chinese heterogeneous encyclopedia knowledge base, *Journal of computational and applied*, 7(36), 1881–1886, 2016.

[18] Zhang, C.; Chen, L.; Li, Q. (2016). Chinese text similarity algorithm based on PST_LDA, *Application Research of Computers*, 33(2), 375–377,383, 2016.

[19] Deng, Y.; Chen, L.; Wu, Y.; et al. (2023). Event entity alignment for multi-source encyclopedia knowledge bases with the similarity of event element sets, *Third International Seminar on Artificial Intelligence, Networking, and Information Technology (AINIT 2022)*, SPIE, 12587, 144–151, 2023.

[20] Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data, *Advances in neural information processing systems 26*, 2013.

[21] Yoon, S.; Ko, S.; Kim, T.; et al. (2025). Unsupervised Robust Cross-Lingual Entity Alignment via Neighbor Triple Matching with Entity and Relation Texts, *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, 184–193, 2025.

[22] Kolyvakis, P.; Kalousis, A.; Kiritsis, D. (2018). Deepalignment: Unsupervised ontology matching with refined word vectors, *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 1-6 June 2018*, 2018.

[23] Shahbazi, H.; Fern, X.Z.; Ghaeini, R.; et al. (2019). Entity-aware elmo: Learning contextual entity representation for entity disambiguation, *arxiv preprint arxiv:1908.05762*, 2019.

[24] Wu, Y.; Liu, X.; Feng, Y.; et al. (2019). Relation-aware entity alignment for heterogeneous knowledge graphs, *arxiv preprint arxiv:1908.08210*, 2019.

[25] Zhu, Y.; Liu, H.; Wu, Z.; et al. (2021). Relation-aware neighborhood matching model for entity alignment, *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5), 4749–4756, 2021.

[26] Bai, L.; Song, C.; Zhu, L. (2024). Joint multi-feature information entity alignment for cross-lingual temporal knowledge graph with bert, *IEEE Transactions on Big Data*, 11(2), 345–358, 2024.

[27] Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach, *arXiv preprint arXiv:1907.11692*, 2019.

[28] Zhang, Z.; Han, X.; Liu, Z.; Jiang, X.; Sun, M.; Liu, Q. (2019). ERNIE: Enhanced Language Representation with Informative Entities, *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, July 28–August 2, 2019, pp. 1441–1451, 2019.

[29] Yang, Z.; Dai, Z.; Yang, Y.; Carbonell, J.; Salakhutdinov, R.; Le, Q. V. (2019). XLNet: Generalized Autoregressive Pretraining for Language Understanding, *Advances in Neural Information Processing Systems*, Vol. 32, pp. 5753–5763, 2019.

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Min Zhang, Luya Yang, Yaxian Gao, Yong Ren (2026). A joint-embedding framework fusing multi-feature information for cultural relic entity alignment in knowledge graphs, *International Journal of Computers Communications & Control*, 21(4), 7233, 2026.
https://doi.org/10.15837/ijccc.2026.4.7233