Research on toponym Knowledge Graph Entity Alignment Based on Spatio-temporal Heterogeneous-aware Attention Mechanism

Han Zhou *, Chenchen Luo, Jiachen Bo, Liangjun Gao

College of Civil Engineering, Hefei University of Technology, Hefei, China, 230009

* Corresponding Author Email: m15256202940@163.com

Abstract. Toponym knowledge graphs (PNKGs) are pivotal in applications such as smart cities, yet the heterogeneity of multi-source data significantly hampers the integration and utility of geographic knowledge. Additionally, few existing approaches account for temporal dynamics in PNKGs. To address these limitations, this study proposes STGEA, an entity alignment method that integrates spatial and temporal relationships within toponym knowledge graphs. Initially, a pre-trained model generates semantic-aware embeddings to capture cross-graph entity similarities, then this study designs a heterogeneous graph transformer equipped with temporal relationship awareness, this transformer employs an attention mechanism to integrate spatial, temporal, and attribute information across toponyms. Experimental results show that STGEA outperforms state-of-the-art methods across key metrics: Hits@1 (94.8%, +3.6%), Hits@5 (98.1%, +2.7%), and MRR (0.96, +0.3). Ablation studies further demonstrate the critical role of spatial and temporal relationships in entity alignment tasks, with performance degradation observed when either relationship is removed (Hits@1 drops to 90.2% and 90.8%, respectively).

Keywords: Entity alignment, Toponym knowledge Graph, Spatiotemporal Relationships, Representation Learning.

1. Introduction

The rapid development of Geographic Information Systems (GIS) and the accumulation of geographic data have positioned place name knowledge graphs as a critical form of knowledge representation. These graphs are pivotal in applications such as smart cities, geographic information services, and cultural heritage preservation. However, the heterogeneity of data sources and the variability in data collection standards often result in inconsistencies in both data representation and integration, hindering the effective utilization of geographic knowledge.

Entity alignment (EA) seeks to establish correspondences between identical entities across different knowledge graphs, enabling the transfer of valuable knowledge from one graph to another. This not only enriches the content of both graphs but also enhances the performance of downstream applications, such as recommendation systems and data integration [1].

Entity alignment techniques can be broadly categorized into two approaches: translation-based methods and graph neural network GNN-based methods [2]. Translation-based models, such as TransE [3] and its extensions (MTransE, IPTransE), embed entities and relations into a low-dimensional vector space, bringing similar entities closer together. While effective for simple relationships, these models struggle with complex, heterogeneous relations. In contrast, GNN-based methods, such as GCN-Align and AttrGNN [4], leverage neighborhood and attribute information to enhance entity alignment performance. However, these methods often fail to address complex, nonlinear heterogeneous relationships. RDGCN [5] introduces relational domains to mitigate structural differences between graphs, while RPR-RHGT [6] integrates relational and path structure information to handle heterogeneity. Yet, these models neglect attribute information, which is vital for accurate entity matching.

Despite the advancements in these methods, few have adequately incorporated temporal information, which is crucial for aligning dynamically evolving place name entities. TEA-GNN [7] is the first method to introduce temporal data into entity alignment, but it fails to incorporate attribute information and struggles with handling complex, multi-relational data. Similarly, DualMatch [8] frames entity alignment as a graph-matching problem, combining relational and temporal features to improve alignment accuracy. However, this approach significantly increases computational complexity, reducing efficiency.

Despite progress in translation-based and GNN-based methods, several challenges remain. Temporal information is often overlooked, particularly for evolving place names, as most methods focus on static features. Rule-based methods also struggle with scalability and heterogeneous geographic data, while existing approaches fail to effectively aggregate or convey the diversity of place name attributes. To address these issues, this study proposes STGEA, a novel entity alignment model that integrates spatiotemporal dynamics and attribute heterogeneity. The model uses pre-trained embeddings to capture semantic information, applies multi-level linear mappings to extract features, and leverages attention mechanisms to capture spatiotemporal interactions. Entity similarity is computed using Manhattan distance, and optimization is achieved with a margin-based ranking loss function. STGEA effectively addresses the temporal and attribute challenges, enhancing entity alignment across time and domains.

2. Ontology Construction and Element Framework for Toponym Knowledge

BP neural network is a multi-layer network with error reverse propagation, which is composed of input layer nodes, hidden layer nodes, and output layer nodes. This process has been reduced to an acceptable level of error to the network output, or to a predetermined number of learning times. The network structure is shown in Figure 1. Place names, as linguistic symbols, denote geographic objects and phenomena with specific locations, extents, and characteristics. They are fundamental to geographic information systems (GIS) and public knowledge, playing a key role in governance, economic development, and cultural heritage. Despite the proliferation of place name databases, their limited attribute information restricts their representation of geographic knowledge. To address this, the study proposes a novel ontology construction method for place name knowledge, culminating in a comprehensive knowledge graph. Unlike traditional databases, this knowledge graph captures complex relationships—such as hierarchical and spatial links—enhancing semantic depth and practical utility.

The construction process begins with a conceptual model based on a top-down, ontology-driven approach. A schema layer standardizes entities, attributes, and relationships, ensuring semantic coherence [9]. Classification and attributes, such as population, coordinates, and historical events, are defined using geographic science principles, while logical relationships like territorial and spatiotemporal links are established. Data from diverse sources—administrative records, mapping services, and historical documents—are processed with natural language processing (NLP) and deep learning techniques, and instantiated into graph structures stored in databases like Neo4j.

The knowledge system is organized into four elements: name, classification, spatiotemporal, and attribute. Name elements include identifiers such as standard forms, aliases, and linguistic variations. Classification elements categorize place names by function, geography, or administrative levels. Spatiotemporal elements cover changes over time and spatial relationships, while attribute elements provide textual and numerical data, such as cultural insights or population size. Together, these elements transform place names into rich, multi-dimensional data points.

This framework offers a comprehensive representation of place name knowledge, enabling applications across domains such as urban planning, environmental studies, and historical research.

3. Spatio-temporal Feature Driven Toponym Entity Alignment Model

3.1. Problem Definitions

Knowledge graphs (KGs) represent entities and concepts as nodes and their relationships as edges. The task of place name entity alignment seeks to match entities in different KGs that represent the same real-world object, thereby obtaining more complete factual knowledge. In this paper, the term place name entity refers specifically to entities within place name knowledge graphs, as shown in Fig 1.

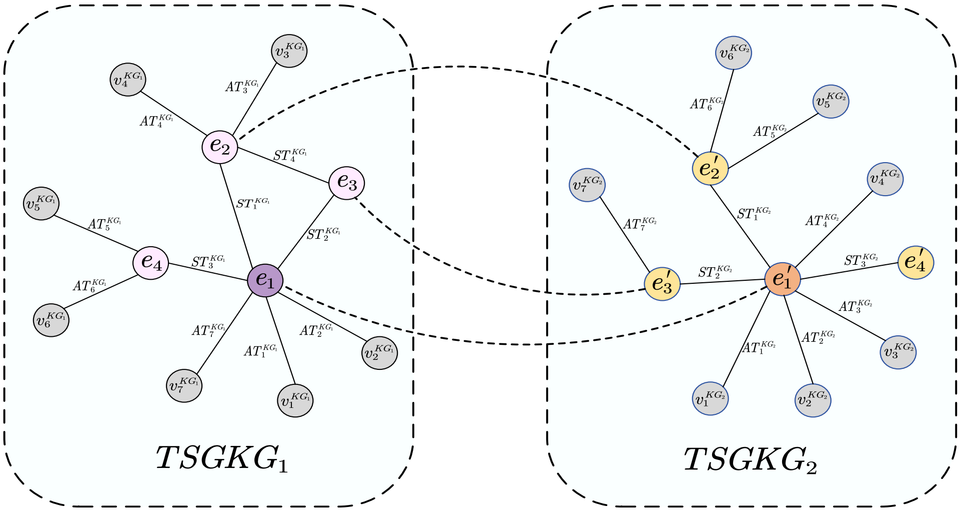

Figure 1. Place Name Knowledge Graph Alignment Task.

In Figure 1, e_1 and e_1', e_2 and e_2', e_3 and e_3' represent the entities to be aligned. The attribute values $v_1^{KG_1}$ and $v_2^{KG_1}$, as well as $v_1^{KG_2}$ and $v_2^{KG_2}$, correspond to the respective values in knowledge graphs KG_1 and KG_2. Entity relationships in the knowledge graphs denoted as $ST_1^{KG_1}$ and $ST_2^{KG_2}$, include temporal relationships (τ) and spatial relationships (s). Similarly, attributes $AT_1^{KG_1}$ and incorporate temporal relationships (τ) and additional attributes (a).

3.2. Overview of STGEA

In Section 3.1, the task of spatiotemporal place name knowledge graph entity alignment was formally defined. This section provides an in-depth description of the proposed STGEA model, designed to align entities within a spatiotemporal place name knowledge graph ($TSGKGs$), as illustrated in Figure 2. The model leverages the importance of spatiotemporal information and addresses the heterogeneity of entity neighborhoods through an attention-based approach. Given two spatiotemporal place name knowledge graphs, $G_1 = (E_1, V_1, ST_1, AT_1)$ and $G_2 = (E_2, V_2, ST_2, AT_2)$, the STGEA framework consists of three core components: the embedding learning module, the aggregation module, and the alignment module.

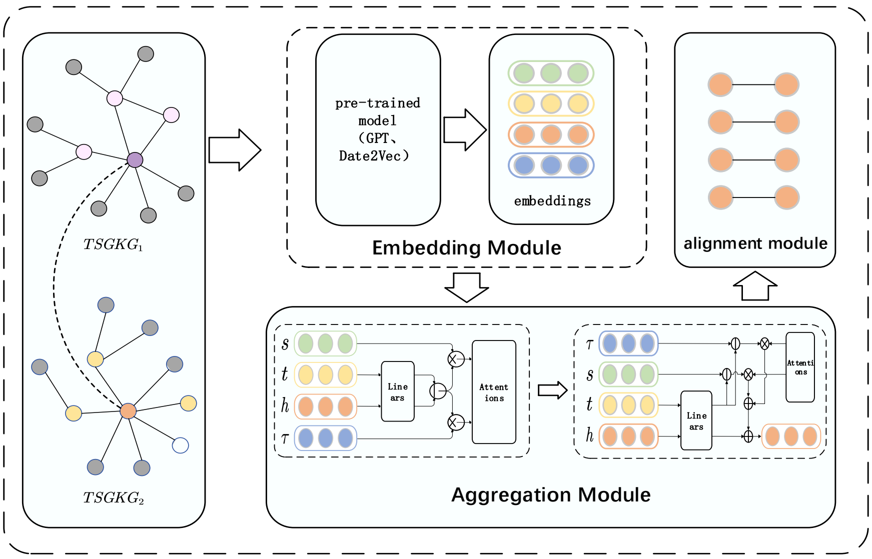

Figure 2. STGEA Framework.

3.3. Embedding Module

When embedding knowledge graphs with rich semantic information, leveraging pre-trained models can significantly enhance model performance [10]. In this study, the pre-trained models $text-embedding-3-small$ and $Date2Vec$ are utilized to generate embedding vectors, as illustrated in Figure 3.

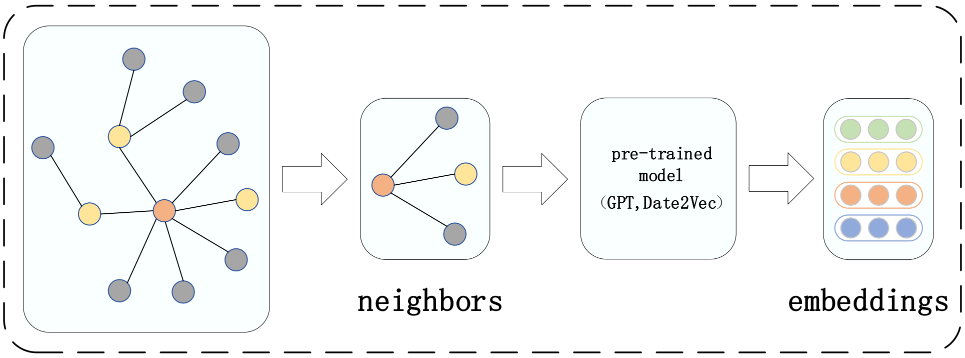

Figure 3. Embedding Module.

Place names, as critical geographic information, exhibit diverse and complex attributes. These attributes include numerical values, such as "population" and "area," which reflect socioeconomic characteristics and geographic scale, as well as textual descriptors, such as "place name origin" and "place name culture," which convey cultural and historical contexts. Capturing the semantic meaning of attribute values remains a significant challenge. Traditional methods based on string similarity often fail to address heterogeneity issues, including synonyms, polysemy, abbreviations, and variations in precision and units. Furthermore, structural discrepancies between different knowledge graphs complicate the alignment task.

To address these challenges, AttrGNN proposed dividing knowledge graphs into subgraphs based on attribute types—such as names, text, numerical, and non-attributes—and employed a BERT encoder to capture the contextual relationships among attribute values. Building on this foundation, the present study utilizes a GPT pre-trained model $text-embedding-3-small$ to generate embedding representations. By leveraging extensive external knowledge and advanced contextual understanding capabilities, the model captures fine-grained semantic relationships between attribute values, enabling unified mapping of heterogeneous attributes. This approach significantly enhances the expressiveness of place name knowledge graphs and provides precise attribute representations, thereby improving the accuracy of geographic information analysis tasks.

Time represents a critical dimension in spatiotemporal place name knowledge graphs. To effectively capture temporal features, this study employs the $Date2Vec$ encoder for temporal embedding. $Date2Vec$ Utilizes an Autoencoder (AE) architecture comprising an encoder and a decoder. The encoder extracts temporal features, while the decoder reconstructs the original time values. By minimizing the reconstruction loss, the decoded time values are aligned as closely as possible with the original inputs. Once the model is trained, pre-trained weights are obtained, and the encoder module is employed to extract temporal features, yielding temporal embedding vectors. In the decoder, the temporal feature vector is processed through two fully connected layers to reconstruct the original time values. The reconstruction accuracy is optimized by minimizing the mean squared error (MSE) loss between the actual and reconstructed time values, ensuring close approximation to the real inputs, as illustrated in Equation (3).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{1}$$

In this context, y_i denotes the element at a position i in the real-time value vector, and $\hat{y}_i$ refers to the corresponding element at position i in the reconstructed time value vector.

3.4. Aggregation Module

3.4.1. Spatiotemporal Heterogeneous-Aware Attention Calculation

Entity alignment tasks are often complicated by the heterogeneity and noise introduced by differing data sources, making direct entity matching a significant challenge. To improve precision and robustness, this study incorporates an attention mechanism that dynamically assigns weights based on the relevance of different components in the input data. This mechanism enables the model to emphasize critical information while disregarding irrelevant noise. To address this, a mapping step is introduced to enhance the representational capacity of each entity, accommodating its multiple meanings under different relational contexts. To underscore the importance of spatiotemporal relationships in place name entity alignment, the model avoids directly computing attention scores via a simple dot product of head and tail entities. Instead, the head entity h and tail entity t are concatenated, followed by a dot product operation with spatiotemporal relationships. This approach effectively captures the interaction between head and tail entities in the context of spatiotemporal relationships, as illustrated in Figure 4.

In this mechanism, the head entity h is mapped to a key vector $K_i(h)$, while the tail entity t is mapped to a query vector $Q_i(t)$. The spatial relationship-weighted importance $H^{attention}(h,s,t)$ and temporal relationship-weighted importance $H^{attention}(h,\tau,t)$ for each tail entity are defined as follows:

$$H_i^{attention}(h,s,t) = \frac{a^T([K_i(h) \parallel Q_i(t)]S^l(s))}{\sqrt{d/h_n}} \tag{2}$$

$$H_i^{attention}(h,\tau,t) = \frac{a^T([K_i(h) \parallel Q_i(t)]T^l(\tau))}{\sqrt{d/h_n}} \tag{3}$$

According to the definition of Graph Attention Network (GAT, 2018), this study employs the softmax function to normalize elements that represent temporal and relational connectivity. This normalization ensures that the attention weights assigned to neighboring entities sum to 1, enabling a balanced and interpretable weighting mechanism:

$$H^{attention}(h,s,t) = \parallel_{i\in[1,h_n]\forall(a,v)\in AN(h)} Soft\max(H_i^{attention}(h,s,t)) \tag{4}$$

$$H^{attention}(h,\tau,t) = \parallel_{i\in[1,h_n]\forall(a,v)\in AN(h)} Soft\max(H_i^{attention}(h,\tau,t)) \tag{5}$$

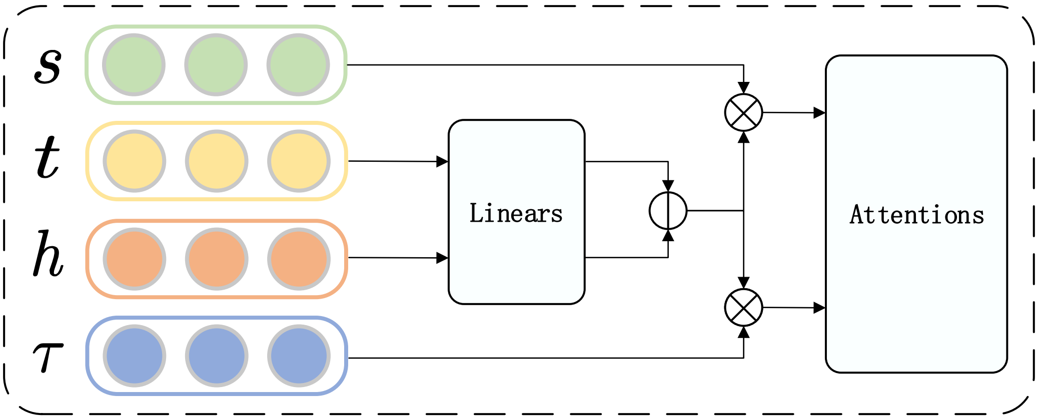

Figure 4. Spatiotemporal Heterogeneous-Aware Attention Calculation

3.4.2. Message Passing and Aggregation

The process of message passing and aggregation is illustrated in Figure 5. Similar to the calculation of attention scores, to fully capture the complex heterogeneity of neighborhood information, the tail entity is first subjected to multiple mappings during the aggregation process. This approach ensures that neighborhood features are adequately expressed. The mapped features are then concatenated with spatiotemporal relationship features, reflecting the multiple meanings of entities under different relationships.

The message-passing mechanism combines the features of the tail entity t, the spatial relationship s, and the temporal relationship τ, transmitting information in a multi-head format. The messages for each head, $H_i^{message}(h,s,t)$ and $H_i^{message}(h,\tau,t)$, are defined as follows:

$$H_i^{message}(h,s,t) = \left[V_{Linear_i}\left(t^{(l-1)}\right) \parallel S^l(s)\right] \tag{6}$$

$$H_i^{message}(h,\tau,t) = \left[V_{Linear_i}\left(t^{(l-1)}\right) \parallel T^l(\tau)\right] \tag{7}$$

Then, all the heads are aggregated to form the final multi-head message:

$$H^{message}(h,s,t) = \parallel_{i\in[1,h_n]} H_i^{message}(h,s,t) \tag{8}$$

$$H^{message}(h,s,t) = \parallel_{i\in[1,h_n]} H_i^{message}(h,\tau,t) \tag{9}$$

The aggregated neighbor messages are weighted by attention scores, integrating information from all neighbors to update the entity embeddings. The aggregation process is as follows:

$$\tilde{e}_h^{(l)} = \sigma\left(\sum_{\forall (a,v)\in AN(h)} \begin{matrix} H^{attention}(h,s,t)\cdot H^{message}(h,s,t) \\ +H^{attention}(h,\tau,t)\cdot H^{message}(h,\tau,t) \end{matrix}\right) \tag{10}$$

Finally, the entity embeddings are updated through residual connections:

$$e_h^{(l)} = \omega A_{Linear\left(\tilde{e}_h^{(l)}\right)} + (1-\omega) N_{Linear\left(e_h^{(l-1)}\right)} \tag{11}$$

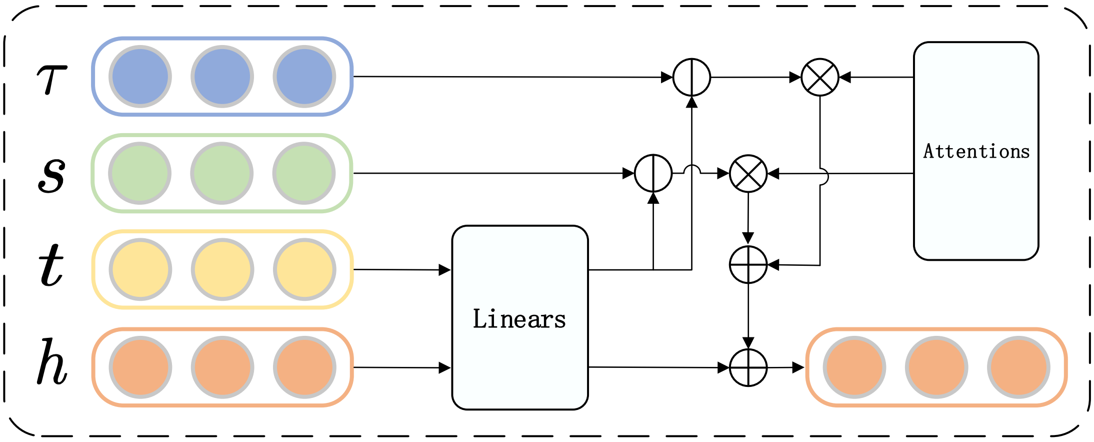

Figure 5. Message Passing and Aggregation.

3.5. Alignment module

The fundamental idea of STGEA is to embed entities from different knowledge graphs into a unified vector space, ensuring that equivalent entities are as close as possible. Aligned entities are subsequently identified by measuring the distances between entities in the embedding space. To quantify the similarity between potential entity pairs, this study employs the Manhattan distance:

$$d_f(e_{1i}, e_{1j}) = \|e_{1f,i} - e_{2f,i}\|_{L1} \tag{12}$$

Here, $f \in \{rel, attr\}$ represents the relation embedding and attribute embedding, and $L1$ denotes the Manhattan distance. However, this study posits that relation embeddings and attribute embeddings may contribute differently to the entity alignment (EA) task, as the structural differences between these two types of embeddings can be significant. Therefore, instead of directly concatenating these embeddings, different weights are assigned to distinguish their contributions during the training process.

To achieve this, a margin-based ranking loss function is employed during training. This function aims to minimize the embedding distance for positive sample pairs while maximizing the distance for negative sample pairs. The specific loss function is defined as follows:

$$\begin{aligned} L = & \sum_{(p,q) \in D, (p',q') \in D'_r} max[0, d_r(p,q) - d_r(p',q') + \gamma_1] \\ & + \beta \sum_{(p,q) \in D, (p',q') \in D'_a} max[0, d_a(p,q) - d_a(p',q') + \gamma_2] \end{aligned} \tag{13}$$

4. Experimental Assessment of Spatio-temporal Toponym Entity Alignment

4.1. Datasets

Guided by the toponymic knowledge graph schema, this study extracts toponym data from numerous toponymic databases, such as GeoNames, GEONet Names Server, OpenStreetMap, Tianditu, and the China Toponym Information Database, as well as from map service providers like Baidu Maps, Amap, and online encyclopedias such as Baidu Baike, Wikipedia, local gazetteers, and government announcements. After manual preprocessing, two isomorphic toponymic knowledge graphs were constructed. The dataset includes a total of 42,931 toponym entities, 215,436 entity triples, 328,448 attribute triples, and 138,743 temporal triples.

4.2. Parameter Settings

This paper selects three commonly used entity alignment evaluation metrics: Hits@1, Hits@10, and MRR. Hits@m indicates the proportion of valid entity pairs for all target entities, where the correct aligned entity appears among the top m candidate entities. MRR is the average reciprocal rank of the correct entity. To ensure fairness in evaluation, this study employs a five-fold cross-validation method. Specifically, the dataset is divided into five mutually exclusive parts, with each part representing 20% of the total dataset. In each experiment, one part is selected as the training data, 10% of the remaining data is used for validation, and the remaining 70% is used for testing. The hyperparameter settings are as follows: maximum training epochs: 2000; batch size: 5000; embedding dimension: 100; margin: 1.5; learning rate: 0.001.

4.3. Main Results and Analysis

4.3.1. Comparative Experiment

To validate the performance of the STGEA model, this paper selects several representative baseline models and compares them using the Hits@1, Hits@10, and MRR metrics. The results, as shown in Table 1, demonstrate that the STGEA model has a significant advantage in handling spatiotemporal relationships and heterogeneous attributes.

Table 1. The results of the comparative experiment.

Model	Hits@1	Hits@5	MRR
MTransE	24.0	43.6	0.34
GCN-Align	49.3	80.4	0.61
HGCN	74.0	87.7	0.78
RDGCN	74.8	85.5	0.76
NMN	76.8	89.2	0.80
AttrGNN	87.4	95.1	0.90
BERT-INT	91.2	95.4	0.93
SSARLM	83.2	90.5	0.87
TEA-GNN	79.6	90.2	0.83
STGEA	94.8	98.1	0.96

Among these baseline models, MTransE mainly relies on the triple information in the knowledge graph, focusing on learning the relationship vectors while neglecting other rich entity attribute information. Some well-performing graph neural network (GNN)-based models, such as RDGCN, HGCN, and NMN, integrate entity neighbor or relation information into the entity embeddings. However, they still fail to consider entity attributes and do not optimize the similarity matrix. While GCN-Align and AttrGNN take entity attribute information into account to some extent, GCN-Align's design is overly simplistic, leading to insufficient alignment accuracy when handling complex entity attributes. AttrGNN, on the other hand, does not consider the spatiotemporal relationships of place names, where changes in spatiotemporal factors may introduce errors in aggregated information, thus reducing alignment accuracy.

Although BERT-INT considers structural information, attribute information, and entity names or descriptions, it overlooks temporal relationship information. TEA-GNN incorporates temporal information but ignores a significant amount of attribute data. In contrast, the STGEA model comprehensively considers four dimensions of information: structure, names, attributes, and spatiotemporal relationships. This multi-dimensional approach allows STGEA to outperform several existing baseline models in terms of alignment performance, especially in complex toponym entity alignment tasks, where it demonstrates stronger adaptability and accuracy.

Furthermore, the STGEA model's text encoder uses a text embedding generator based on a pre-trained model, which is trained on a large-scale dataset encompassing rich textual data, including content in multiple languages. This enables the model to have strong comprehension capabilities, capturing rich semantic information and contextual relationships. The model strikes a good balance between accuracy and computational speed, effectively generating high-quality vector representations, thereby further enhancing overall model performance.

4.3.2. Ablation Study

Table 2 presents the performance comparison of different variants of the STGEA model in an ablation study, aiming to validate the impact of spatial and temporal relationships on toponym entity alignment performance.

Table 2. The results of the ablation study.

Model	Hits@1	Hits@5	MRR
STGEA	94.8	98.1	0.96
STGEA(W/O S)	90.2	95.7	0.92
STGEA(W/O T)	89.3	93.9	0.91
STGEA(W/O ST)	82.6	90.8	0.88

STGEA (W/O S) refers to the ablation experiment that excludes spatial relationships, using only latitude and longitude coordinates as spatial information for entity alignment. STGEA (W/O T) represents the ablation experiment that disregards temporal relationships, considering only relationships and attribute triples without accounting for temporality. STGEA (W/O ST) combines both omissions, removing spatial and temporal relationships entirely. These ablation studies were conducted to evaluate the impact of spatial and temporal relationships on the performance of toponym entity alignment.

The experimental results demonstrate that spatial relationships significantly enhance the model's ability to capture complex spatial correlations between neighboring toponyms, exerting a notable influence on the attention aggregation module. Temporal relationships, on the other hand, are critical for addressing changes in toponym relationships and attribute information. The temporal-aware module effectively mitigates alignment accuracy degradation caused by evolving information. Furthermore, the results reveal that excluding spatial or temporal relationships leads to a measurable decline in the Hits@m and MRR metrics, underscoring the indispensable role of integrating both spatial and temporal relationships in addressing complex toponym entity alignment tasks.

5. Conclusions

Toponym entity alignment is intrinsically linked to spatiotemporal relationships, while also encompassing complex attribute features and heterogeneous structural information, making the task particularly challenging. This study proposes STGEA, a spatiotemporal relationship-enhanced model specifically designed to address the complexities of toponym entity alignment. STGEA incorporates attribute value information using a pre-trained language model and integrates heterogeneous relationships and attributes through an attention mechanism, resulting in more precise embedding representations. By fully leveraging spatiotemporal relationships, the model enhances information aggregation accuracy and significantly improves alignment outcomes. In addition to exploring the spatiotemporal dynamics between toponym entities and their attributes, STGEA addresses the heterogeneity of relationships and attributes, integrating relational structures, attribute structures, and spatiotemporal information to achieve semantic enhancement. This approach improves the precision and robustness of alignment, with STGEA demonstrating superior performance compared to state-of-the-art methods, particularly excelling in the Hits@1 metric.

Despite these advancements, STGEA faces limitations in handling numerical attributes, particularly in knowledge graphs rich in numerical data. This challenge highlights the need for pre-trained models tailored to the toponym domain, which can effectively integrate both textual and numerical attribute values. Future work will focus on developing such domain-specific models and optimizing the architectural design of STGEA to further harness spatiotemporal information and maximize its potential.

This study reveals that leveraging spatiotemporal relationships and heterogeneous attribute information can help solve the challenges of toponym entity alignment, particularly in handling complex features and structural heterogeneity.

References

[1] Zhu B, Wang R, Wang J, et al. A survey: knowledge graph entity alignment research based on graph embedding [J]. Artificial Intelligence Review, 2024, 57(9): 229.

[2] Lu D, Han G, Zhao Y, et al. Review of Deep Learning-Based Entity Alignment Methods[C]//International Conference on Green, Pervasive, and Cloud Computing. Singapore: Springer Nature Singapore, 2023: 61-71.

[3] Bordes A, Usunier N, Garcia-Duran A, et al. Translating embeddings for modeling multi-relational data[J]. Advances in neural information processing systems, 2013, 26.

[4] Liu Z, Cao Y, Pan L, et al. Exploring and evaluating attributes, values, and structures for entity alignment [J]. arXiv preprint arXiv:2010.03249, 2020.

[5] Wu Y, Liu X, Feng Y, et al. Relation-aware entity alignment for heterogeneous knowledge graphs [J]. arXiv preprint arXiv:1908.08210, 2019.

[6] Cai W, Ma W, Zhan J, et al. Entity alignment with reliable path reasoning and relation-aware heterogeneous graph transformer [J]. arXiv preprint arXiv:2205.08806, 2022.

[7] Xu C, Su F, Lehmann J. Time-aware graph neural networks for entity alignment between temporal knowledge graphs. arXiv 2022[J]. arXiv preprint arXiv:2203.02150.

[8] Liu X, Wu J, Li T, et al. Unsupervised entity alignment for temporal knowledge graphs[C]//Proceedings of the ACM Web Conference 2023. 2023: 2528-2538.

[9] Zhu Y, Sun K, Wang S, et al. An adaptive representation model for geoscience knowledge graphs considering complex spatiotemporal features and relationships [J]. Science China Earth Sciences, 2023, 66(11): 2563-2578.

[10] He L, Li H, Zhang R. A Semantic-Spatial Aware Data Conflation Approach for Place Knowledge Graphs [J]. ISPRS International Journal of Geo-Information, 2024, 13(4): 106.