EAI Endorsed Transactions
on Scalable Information System

Research Article EAI.EU

Leveraging Relation Attention Mechanisms for Enhanced Knowledge Graph Completion with Embedding Translation

Jiahao Shi[1], Zhengping Lin[1,*], Yuzhong Zhou[1], Yuliang Yang[1], and Jie Lin[1]

[1]Electric Power Research Institute of China Southern Power Grid Company, Guangzhou, 510663, China

Abstract

In this paper, we propose a novel knowledge graph completion framework to leverage a relation-specific attention mechanism integrated with an embedding translation strategy to improve the accuracy and contextual understanding of link prediction tasks. Unlike traditional models that rely on fixed transformation spaces, the proposed method dynamically captures fine-grained relational semantics by combining hierarchical candidate categorization, relation-guided entity projection, and asymmetric score functions. Specifically, the proposed model employs K-means clustering and principal component analysis (PCA) to identify semantically consistent entity sets, and integrates attention-weighted multi-attribute embeddings to construct robust relational representations. A margin-based ranking loss with normalized embedding constraints ensures effective optimization, further supported by Xavier initialization and stochastic gradient descent. Extensive experiments on two benchmark datasets, WN18 and FB15K, demonstrate the superiority of the proposed method. Specifically, on WN18, the proposed method achieves the lowest mean rank (MR) of 144, with competitive results in mean reciprocal rank (MRR) (0.902), Hits@1 (89.0%), Hits@3 (90.4%), and Hits@10 (96.3%), closely rivaling state-of-the-art models like QuatE and ComplEx. On FB15K, the proposed method again delivers the best Mean Rank of 21, along with strong scores in MRR (0.831), Hits@1 (72.2%), Hits@3 (88.4%), and the highest Hits@10 (92.5%) among all compared methods.

Received on 18 April 2025; accepted on 19 September 2025; published on 07 October 2025

Keywords: Knowledge graph, relation attention mechanism, embedding translation, performance evaluation.

doi:10.4108/eetsis.9117

1. Introduction

The rapid advancement of information technology (IT) has significantly transformed the landscape of industrial systems, leading to the emergence of the industrial Internet of Things (IIoT), a paradigm that integrates traditional manufacturing processes with cutting-edge digital connectivity and data analytics [1–3]. IIoT networks leverage a multitude of interconnected devices-ranging from sensors, actuators, and edge computing nodes to cloud-based platforms-to facilitate real-time monitoring, control, and optimization of industrial operations. Unlike conventional IT systems, IIoT emphasizes ultra-reliable, low-latency communication, high scalability, and strong interoperability across heterogeneous devices and protocols, all while adhering to stringent security and privacy requirements [4, 5]. Some key enabling technologies, such as 5G and beyond, software-defined networking (SDN), edge/fog computing, and AI-driven analytics, have been proposed to collectively enhance the intelligence, flexibility, and responsiveness of industrial systems [6, 7]. In addition, time-sensitive networking (TSN) and deterministic communication protocols have been investigated to support mission-critical applications in sectors such as smart manufacturing, energy, and logistics. Moreover, distributed ledger technologies like blockchain have been explored to ensure data integrity and trustworthiness across decentralized IIoT ecosystems [8, 9].

*Corresponding author. Email: zhengping_lin@hotmail.com

Knowledge graphs (KGs) have emerged as a powerful paradigm for structuring and representing semantic relationships among entities, and their integration into IIoT networks has attracted growing attention in recent research [10–12]. In the context of IIoT, knowledge graphs offer a semantic layer that enables intelligent data integration, contextual reasoning, and dynamic decision-making by capturing complex relationships between machines, processes, sensors, and operational events [13, 14]. The role of KGs in has been investigated in enhancing interoperability across heterogeneous IIoT devices and systems by providing a unified ontology-driven representation that bridges syntactic and semantic gaps [15]. Domain-specific ontologies, such as SAREF and SSN/SOSA, have been proposed to model manufacturing knowledge, asset hierarchies, and operational workflows [16, 17]. Moreover, the application of graph neural networks (GNNs) and knowledge graph embeddings has further advanced predictive maintenance, anomaly detection, and supply chain optimization within industrial environments. Moreover, the fusion of KGs has been explored with edge computing and AI techniques, in order to enable real-time reasoning at the edge, improving responsiveness and scalability [18].

The attention mechanism has become a pivotal component in advancing intelligent analytics and decision-making within IIoT networks, where massive, heterogeneous, and dynamically evolving data streams present significant challenges to traditional modeling approaches [19, 20]. In IIoT applications, attention mechanisms enable models to selectively focus on the most relevant features, time steps, or nodes in a network, thus enhancing the interpretability, efficiency, and performance of learning systems [21, 22]. Temporal attention has been exploited to capture long-range dependencies in time-series sensor data for applications such as predictive maintenance and equipment failure forecasting. Meanwhile, spatial attention has been employed to dynamically prioritize inputs from critical machines or processes in multi-source monitoring environments [23, 24]. The integration of attention with graph neural networks, particularly in the form of graph attention networks (GATs), has shown effectiveness in modeling complex inter-device and inter-process relationships, supporting tasks like anomaly detection, industrial knowledge graph completion, and cyber-physical system state estimation. Moreover, transformer-based architectures, originally designed for language processing, have been adapted for spatio-temporal forecasting and event recognition in smart factories, offering scalability and robust performance across varying industrial settings [25]. At the network edge, lightweight attention modules can be embedded in edge AI systems to enable context-aware inference with reduced computational overhead, which is crucial for latency-sensitive IIoT applications [26–28].

Motivated by the above literature review, this paper introduces a novel framework for knowledge graph completion that integrates a relation-aware attention mechanism with a dynamic embedding translation strategy to enhance the precision and contextual relevance of link prediction. In contrast to conventional embedding models that operate within static transformation spaces, the proposed method adaptively models fine-grained relational semantics by incorporating hierarchical candidate filtering, relation-specific projection, and asymmetric scoring functions. The framework employs K-means clustering to preselect semantically coherent candidate entities and utilizes principal component analysis (PCA) to identify the most informative dimensions for each relation. An attention-based multi-attribute embedding scheme is then applied to capture nuanced entity features under varying relational contexts. Training is guided by a margin-based ranking loss combined with embedding normalization constraints, and the model is optimized using Xavier initialization and stochastic gradient descent to ensure stable and efficient convergence. Finally, extensive empirical evaluations on the widely used WN18 and FB15K datasets validate the effectiveness of the proposed method. Specifically, on WN18, the model attains the lowest mean rank (MR) of 144, along with competitive results in mean reciprocal rank (MRR) (0.902), Hits@1 (89.0%), Hits@3 (90.4%), and Hits@10 (96.3%), performing on par with or surpassing advanced baselines such as QuatE and ComplEx. On FB15K, the proposed method consistently outperforms existing methods, achieving the best MR of 21, a robust MRR of 0.831, and leading values in Hits@1 (72.2%), Hits@3 (88.4%), and Hits@10 (92.5%). These results highlight the proposed method's capability to effectively capture both semantic richness and structural patterns in knowledge graphs.

2. System Model

In the context of human cognition, the process of categorizing relationships follows a hierarchical model, wherein entities are grouped according to shared attributes. These attributes serve as distinguishing features between different categories, while simultaneously helping to differentiate entities within the same category. This hierarchical categorization process is essential in human reasoning, particularly when evaluating relations between entities.

Consider a relational triple of the form $(x, \text{occupational skill} : \text{farming})$, where the entity x needs to be determined. In this context, the cognitive model would exclude entities such as "car" because it does not belong to the category of entities possessing the attribute "occupational skill". This distinction is grounded in the idea that "farming" as an occupational skill can only be meaningfully associated with certain entities, such as a "farmer" or "agricultural worker", but not with a "car". Mathematically, this can be expressed as,

$$\text{If } x \notin \{\text{Farmer, Agricultural Worker}\}, \\ \text{then } x \notin \text{Category(Occupational Skill)}. \tag{1}$$

In this scenario, the relationship is clearly defined by the occupation, leading to an exclusionary process based on categorical attributes. The difference between "farmer" and "driver" is accentuated by their respective skills-while both possess the "occupational skill", only the farmer applies the specific skill of "farming".

Building upon this, the process of categorizing entities for specific relations should extend to fine-grained attribute embedding. The link prediction process for relational tasks occurs in two distinct stages,

- **Candidate Collection**: In this stage, the task involves gathering a set of potential candidate entities from across multiple categories. These entities are pre-filtered based on broad relational attributes.
- **Fine-Grained Categorization**: In this stage, these candidate entities are further analyzed based on more refined attributes to determine the specific relational context. This requires a more granular understanding of each entity's attributes, where embedding models come into play.

Let $R(x, y)$ represent the relation between two entities x and y. The learning objective is,

$$R(x, y) = f(\mathbf{e}_x, \mathbf{e}_y, \mathbf{A}_R), \tag{2}$$

where $\mathbf{e}_x$ and $\mathbf{e}_y$ are the embedding vectors for entities x and y, and $\mathbf{A}_R$ is the transformation matrix that encodes the relation-specific attributes.

To learn the appropriate embedding for these relations, a relation-specific attention mechanism is adopted, which can be mathematically formulated as,

$$\mathbf{e}_R = \sum_i \alpha_i \mathbf{e}_{x_i} + \beta_i \mathbf{e}_{y_i}, \tag{3}$$

where α_i and β_i are the attention weights assigned to different entity components based on their relevance to the target relation R. The embedding model should be trained such that the relation-specific features are effectively captured through the transformation matrices that handle multiple fine-grained attributes across entities.

The primary challenge in relation learning lies in the ability to accurately determine the transformation matrix that can capture the diversity of fine-grained entity characteristics. It is well-established that certain relation models, such as TransH, TransR, and TransE, attempt to learn these transformation matrices. However, these models often fail to capture the multiple attribute dimensions of entities adequately, particularly when the relations between entities are highly nuanced. For instance, as demonstrated by the transformation matrix learned by TransR, it approximates the transformation between entities through a shared embedding space, but it struggles when the fine-grained differentiation of attributes is required. This limitation can be formalized as,

$$\mathbf{T}_{x,y} \approx \mathbf{T}_{x'} \text{ if } \|\mathbf{e}_x - \mathbf{e}_y\|_2 \approx \|\mathbf{e}_{x'} - \mathbf{e}_{y'}\|_2, \tag{4}$$

where $\mathbf{T}_{x,y}$ represents the transformation matrix between entities x and y, and $\mathbf{e}_x$, $\mathbf{e}_y$ are their respective embeddings. Despite the transformation-based models, it is still hard to accurately model the relations, as the fine-grained attributes of entities should be simultaneously considered. In this regard, advanced approaches seek to use multi-dimensional transformation matrices that incorporate the various attributes into the embedding space. The aim is to allow for more accurate categorization, even at the level of highly granular relations. To this end, we can capture the multi-attribute embedding for relation learning, given by

$$\mathbf{e}_{\text{fine-grained}} = \sum_j \mathbf{e}_j(\mathbf{A}_R) \cdot \mathbf{w}_j, \tag{5}$$

where $\mathbf{e}_j$ represents the individual entity embeddings for each attribute, $\mathbf{A}_R$ is the learned transformation matrix, and $\mathbf{w}_j$ are the weights associated with each attribute's contribution to the final relation-specific embedding.

3. Proposed Scheme

Inspired by the previously discussed hierarchical models of cognition, we propose a novel approach for learning embeddings, relation-related candidates, and relation-related attention in a unified manner. This method simultaneously optimizes the entity embeddings while identifying the most relevant candidate entities for a given relation. Our approach is grounded in a piecewise evaluation function that allows for effective evaluation of relational triples. Let $f_r(h, t)$ denote the evaluation function, given by,

$$f_r(h, t) = \begin{cases} P_r(h) + r - P_r(t) & \text{If } h \in H_r, t \in T_r, \\ \infty & \text{Otherwise.} \end{cases} \tag{6}$$

where $P_r(h)$ is the projection of the head entity h related to relation r, $P_r(t)$ is the projection of the tail entity t related to relation r, H_r and T_r represent the set of possible head and tail candidates for the relation r, and r is the relation between the entities. Note that this function is evaluated to infinity if the candidate entities do not belong to the suitable category, revealing that the evaluation function assigns a high cost (infinity) when either the head or the tail entity does not align with the corresponding relation r. If both the head and tail entities are compatible with the relation, their embeddings are evaluated to ensure that their relational distance is minimized, as in the TransE model, while considering only the relevant dimensions.

The difficult now lies in selecting the proper candidates and determining their projection. For this, we adopt the K-means clustering to group entities based on shared characteristics. These clusters are then used to generate candidate sets for a particular relation r, ensuring that the head (tail) candidates are derived from relevant categories. The K-means clustering approach ensures that the model generalizes across different datasets with varying attribute structures.

Next, we examine the structural intricacies of relation learning and propose a method to handle entities with multiple attributes. Based on the principles of PCA, we can know that if a dimension is critical for a specific relation, the variance of the candidates along that dimension will be large. Hence, we leverage a projection function,

$$P_r(h) = a_r \cdot h, \quad a_r \in \mathbb{R}^k, \tag{7}$$

to select the most informative dimensions for each relation r, where $h \in \mathbb{R}^k$ is the entity embedding and $a_r \in \mathbb{R}^k$ is a weight vector for relation r.

Following this approach, we can differentiate entities capable of representing a particular relation r from others. The relation distance is computed based on the selected dimensions, given by,

$$d_r(h,t) = \|P_r(h) - P_r(t)\|_2, \tag{8}$$

where $d_r(h,t)$ represents the distance between the head entity h and the tail entity t under the relation r, and the Euclidean norm $\|\cdot\|_2$ is used to measure the relational distance. This allows to perform fine-grained analysis of the entities that best represent a particular relation.

In cases where some relations exhibit oriented relation chains (ORC) structures, we further extend this model by introducing an asymmetric score function. This function differentiates between the head and tail entities by learning distinct representations for their respective positions in the relation, given by,

$$\|\mathbf{r}\|_r \approx 0. \tag{9}$$

This condition is satisfied by an asymmetric operation applied to the head and tail entity embeddings, resulting in different vector representations for each. Specifically, the head and tail embeddings are scaled by separate transformation matrices, enabling the model to account for their different roles in the relation.

The asymmetric score function is defined as,

$$f_r(h,t) = P_r(\sigma(\mathbf{r}_h h)) + P_r(\sigma(\mathbf{r}_t t)), \tag{10}$$

where P_r is the projection function related to the relation r, σ is the sigmoid function, ensuring that the embedding vectors are scaled appropriately, and $\mathbf{r}_h$ and $\mathbf{r}_t$ are transformation matrices specific to the head and tail entities, respectively. This core function provides an effective method to handle entities with multiple attributes, taking into account both their head and tail representations within the relational context.

To effectively implement the proposed method, we propose the following margin-based ranking loss function for the discrimination of positive and negative triples in knowledge graph completion. This approach ensures that the correct triples (i.e., those in the knowledge graph) are ranked higher than the incorrect (or corrupted) triples during the training process. The loss function L is given by,

$$\begin{aligned} L = &\sum_{h \in H_r, t \in T_r} [f_r(h,t) + \gamma - f_r(h',t')] \\ &+ \alpha \left(\sum_{h \in H_r^+, t \in T_r^+} \left[\|\mathbf{t}' - \mathbf{h}'\|_2^2 + \gamma - \|\mathbf{h} - \mathbf{t}\|_2^2 \right] \right), \end{aligned} \tag{11}$$

where γ is a margin term that enforces a minimum separation between positive and negative triples, h' and t' are corrupted head and tail entities respectively for a given relation r, H_r^+ and T_r^+ denote the positive head and tail entity sets for relation r, and the term $\|\mathbf{h} - \mathbf{t}\|_2$ represents the Euclidean distance between the embeddings of the head and tail entities for a given triple.

The loss function L encourages the model to ensure that the score of the positive triples (h,r,t) is larger than the score of corrupted triples (h',r,t') by a margin γ, while penalizing the model for incorrect triples. The inclusion of α ensures a proper weighting between the various components of the loss function, thus balancing the regularization and ranking objectives. Following prior methods, we enforce constraints on the norms of the entity embeddings h, t, and the relation embeddings r, ensuring that for all triples, the following conditions hold,

$$\|\mathbf{h}\|_2 \leq 1, \quad \|\mathbf{r}\|_2 \leq 1, \quad \|\mathbf{t}\|_2 \leq 1. \tag{12}$$

These constraints indicate that the embeddings of entities and relations remain normalized within a unit sphere in the embedding space, helping to avoid overfitting and ensuring that the embeddings stay well-behaved during the optimization process. The condition on the entity embeddings ensures that both the head and tail entities of a triple are constrained to lie within a normalized space, and similarly, the relation embeddings are also constrained, making it easier for the model to distinguish between valid and invalid triples.

In addition to the margin-based ranking loss, the optimization of the model embeddings also requires an effective sampling strategy. We use the Bernoulli sampling to sample batches of entities and relations during the training process, which aids in the fair comparison of embeddings. To further optimize the embeddings, we utilize Xavier initialization, which initializes the model parameters to prevent issues with vanishing or exploding gradients during training. The Xavier initialization of the entity and relation embeddings is given by,

$$\mathbf{e}_i \sim \mathcal{U}\left(-\sqrt{\frac{6}{N_{\text{in}} + N_{\text{out}}}}, \sqrt{\frac{6}{N_{\text{in}} + N_{\text{out}}}}\right), \tag{13}$$

where $\mathcal{U}$ denotes a uniform distribution, and N_{in} and N_{out} are the number of input and output units (dimensions) in the embeddings. This initialization method ensures that the embeddings start at a reasonable scale, which is crucial for effective gradient propagation during the backpropagation process.

To optimize the embeddings, we apply stochastic gradient descent (SGD), which updates the embeddings iteratively by minimizing the loss function. Specifically, the embeddings are updated by computing the gradient of the loss function with respect to each entity and relation embedding, and then adjusting the embeddings using the gradient, given by,

$$\mathbf{e}_i \leftarrow \mathbf{e}_i - \eta \nabla_{\mathbf{e}_i} L, \tag{14}$$

where $\mathbf{e}_i$ is the embedding of the entity or relation, η is the learning rate, $\nabla_{\mathbf{e}_i} L$ is the gradient of the loss function with respect to the embedding $\mathbf{e}_i$.

By iterating over multiple batches and updating the embeddings accordingly, the model converges to a set of embeddings that are able to accurately predict missing triples in the knowledge graph. Overall, the proposed margin-based ranking loss function, combined with the relation-specific attention mechanism and effective sampling strategies, enables our model to effectively complete knowledge graphs by distinguishing between correct and incorrect triples. By enforcing normalization constraints on the embeddings and using Xavier initialization for stable training, the embeddings can remain effective throughout the learning process. Furthermore, the use of stochastic gradient descent allows for efficient optimization of the model, ensuring that the embeddings improve over time and yield accurate knowledge graph completion. The whole procedure of the proposed scheme in this paper can be summarized in Algorithm 1.

Algorithm 1 Proposed Knowledge Graph Completion Algorithm

Input: Set of entities $\mathcal{E}$, relations $\mathcal{R}$, triples $\{(h, r, t)\}$, relation-specific attention weights α_i, β_i, clustering function $\mathcal{K}$means, PCA function $\mathcal{PCA}$, margin term γ.
Output: Optimized entity and relation embeddings $\{\mathbf{e}_k, \mathbf{r}_k\}$.
Initialization:
Initialize entity embeddings $\mathbf{e}_k$ and relation embeddings $\mathbf{r}_k$ using Xavier initialization.
while not converged **do**
    Candidate Selection:
    Perform $\mathcal{K}$means clustering to select candidate entities for each relation.
    Select the top candidate entities based on relation attributes.
    Fine-Grained Categorization:
    Use $\mathcal{PCA}$ to identify the most informative dimensions for each relation.
    Embedding Update:
    Update the embeddings $\mathbf{e}_k$ using relation-specific attention weights α_i and β_i.
    Compute the projection of head h and tail t entities for each relation r using the projection function $\mathbf{P}_r(h)$ and $\mathbf{P}_r(t)$.
    Calculate the relation distance $d_r(h, t)$ based on the Euclidean norm.
    Loss Calculation:
    Compute the margin-based ranking loss for each relation r using the defined loss function L.
    Update the embeddings using stochastic gradient descent (SGD).
end while
Output: Final optimized entity and relation embeddings.

4. Simulation Results and Discussions

To validate the proposed scheme in this paper, we use the WN18 and FB15K datasets, which are two widely used benchmarks in the field of knowledge graph completion, each containing a rich set of entities and relations that enable comprehensive evaluation of embedding models. Specifically, WN18 is derived from WordNet, a lexical database that categorizes words into synonym sets and represents semantic relationships among them. This dataset includes 18 types of relations and a total of 40,493 entities, covering a variety of semantic relationships such as hypernyms, hyponyms, and meronyms. The dataset contains 141,442 training triples, with 5,000 validation and test triples each, making it suitable for link prediction tasks where the goal is to predict missing entities in a given relation. FB15K, on the other hand, is sourced from Freebase, a large collaborative knowledge base that covers a broad spectrum of domains, from entertainment to politics. It consists of 1,345 types of relations and 14,951 entities, with 483,142 training triples, and 50,000 and 59,071 triples for validation and testing, respectively. Unlike WN18, which focuses on relationships primarily related to word meanings, FB15K contains a wider range of real-world factual relationships, including geographic locations, companies, and academic affiliations. Both datasets serve as excellent testbeds for evaluating models on tasks such as link prediction and triple classification, allowing for a nuanced understanding of how well different algorithms can generalize across different types of knowledge graphs, from semantic lexicons to real-world entities and relationships. These characteristics make both WN18 and FB15K ideal datasets for benchmarking knowledge graph embedding techniques.

Table I presents a comprehensive comparison of link prediction performance across various knowledge graph embedding methods on the WN18 dataset. The table provides the results of five evaluation metrics: MR and MRR, Hits@1, Hits@3, and Hits@10. From this table, we can find that among all methods, the proposed scheme achieves the highest overall MRR (0.982) and top score in Hits@1 (92.1), Hits@3 (97.0), and Hits@10 (95.7), demonstrating its strong performance in capturing semantic relationships in knowledge graphs. However, the proposed method stands out with a significantly superior MR of 144, which is the lowest among all methods£¬ indicating that the proposed method ranks correct entities much closer to the top than any other approach. Additionally, the proposed method delivers competitive performance in MRR (0.902), Hits@1 (89.0), Hits@3 (90.4), and Hits@10 (96.3), achieving a near-optimal balance between the ranking quality and prediction accuracy. Compared to the competing methods such as TransE and DistMult, which either suffer from poor MRR or lower Hits@n, the proposed method consistently provides both accurate and robust prediction. Moreover, even in comparison to state-of-the-art neural models like GIE, R-GCN, and CompGCN, the proposed method exhibits either higher or comparable score, particularly excelling in the ranking accuracy. These results demonstrate the effectiveness of the proposed method, suggesting that it successfully captures the relational and structural dependencies within knowledge graphs, outperforming or matching existing models in both the precision and ranking capability.

Table II provides a detailed performance comparison of several knowledge graph embedding models evaluated on the FB15K dataset for the link prediction task. From this table, we can find that the proposed method achieves the best overall performance in multiple key aspects. Specifically, the proposed method attains the lowest MR of 21, indicating its superior ability to rank the correct tail entities closer to the top of the prediction list. In terms of MRR, the proposed method achieves a high score of 0.831, outperforming other competing methods such as QuatE (0.820), ConvE (0.771), and GIE (0.878). Although GIE achieves slightly higher Hits@1 (78.2%) and Hits@3 (86.7%) compared to the proposed method's 72.2% and 88.4% respectively, the proposed method achieves the highest Hits@10 at 92.5%, highlighting its robustness in top-10 predictions. Compared to the competing models like TransE and DistMult, which yield significantly lower MRRs (0.514 and 0.595 respectively), the proposed method provides a substantial improvement in the predictive precision. Furthermore, even in contrast to more sophisticated models such as R-GCN and CompGCN, the proposed method offers better or comparable performance across all metrics, particularly excelling in balancing both the ranking efficiency and prediction accuracy. The results in Table II further validate that the proposed method effectively captures both semantic and structural dependencies in the knowledge graph, making it a highly effective model for link prediction tasks on the complex datasets.

5. Conclusions

In this paper, we proposed a novel method for knowledge graph completion that integrated a relation-specific attention mechanism with a translation-based embedding strategy to improve the accuracy and contextual understanding of link prediction tasks. Unlike traditional approaches that operated within static transformation spaces, the proposed method dynamically captured fine-grained relational semantics through a combination of hierarchical candidate selection, relation-guided projection, and asymmetric scoring mechanisms. To construct semantically consistent candidate sets, we employed K-means clustering, while PCA was used to identify relation-relevant dimensions with high variance. The attention mechanism weighted multi-attribute entity embeddings according to their relevance to the target relation, enabling more discriminative representation learning. Optimization was achieved using a margin-based ranking loss with embedding norm constraints, supported by Xavier initialization and stochastic gradient descent to ensure stable training. We validated the proposed model on two benchmark datasets, WN18 and FB15K, and the results demonstrated the effectiveness of our approach. On WN18, the proposed method achieved the lowest MR of 144, along with strong performance in MRR (0.902), Hits@1 (89.0%), Hits@3 (90.4%), and Hits@10 (96.3%), closely rivaling or surpassing state-of-the-art methods such as QuatE and ComplEx. On FB15K, the proposed

Table 1. Link prediction performance with various knowledge graph embedding methods: WN18.

Metric	MR	MRR	Hits@1	Hits@3	Hits@10
TransE	-	0.566	15.9	88.2	91.6
RotatE	164	0.926	89.5	95.8	96.3
QuatE	429	0.982	92.1	97.0	95.7
HaKE	336	0.857	93.7	94.9	90.9
GIE	-	0.883	89.4	96.0	95.7
Proposed method	**144**	**0.902**	**89.0**	**90.4**	96.3
DistMult	599	0.796	77.4	87.6	90.4
ComplEx	-	0.973	94.5	96.0	97.0
R-GCN	-	0.737	66.3	87.6	**91.3**
ConvE	536	0.880	90.1	90.6	95.2
ConvKB	556	0.790	78.3	81.0	94.7
CompGCN	575	0.632	55.2	79.5	88.4
NodePiece	689	0.494	28.4	52.9	59.1

Table 2. Link prediction performance comparison with various models: FB15K.

Metric	MR	MRR	Hits@1 (%)	Hits@3 (%)	Hits@10 (%)
TransE	–	0.514	33.2	61.4	71.5
RotatE	32	0.661	62.0	79.2	84.5
QuatE	41	0.820	66.1	81.5	89.1
HaKE	128	0.692	60.0	73.5	78.7
GIE	–	0.878	78.2	86.7	88.1
Proposed method	21	**0.831**	**72.2**	**88.4**	**92.5**
DistMult	42	0.595	50.9	75.1	84.6
ComplEx	–	0.659	55.2	72.4	79.4
R-GCN	–	0.707	59.9	70.1	85.3
ConvE	64	0.771	71.1	82.0	88.2
ConvKB	109	0.609	50.6	67.8	73.0
CompGCN	69	0.382	29.9	47.2	67.7
NodePiece	420	0.167	6.98	17.6	30.4

method again achieved the best MR of 21, with competitive scores in MRR (0.831), Hits@1 (72.2%), Hits@3 (88.4%), and the highest Hits@10 (92.5%).

References

[1] S. Kekki, W. Featherstone, Y. Fang, P. Kuure, A. Li, A. Ranjan, D. Purkayastha, F. Jiangping, D. Frydman, G. Verin *et al.*, "MEC in 5G networks," *ETSI white paper*, vol. 28, no. 28, pp. 1–28, 2018.

[2] S. Guo and X. Zhao, "Multi-agent deep reinforcement learning based transmission latency minimization for delay-sensitive cognitive satellite-UAV networks," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 131–144, 2023.

[3] H. Hui and W. Chen, "Joint scheduling of proactive pushing and on-demand transmission over shared spectrum for profit maximization," *IEEE Trans. Wirel. Commun.*, vol. 22, no. 1, pp. 107–121, 2023.

[4] M. Rahim, G. Kaddoum, and T. N. Do, "Joint devices and IRSs association for terahertz communications in industrial IoT networks," *IEEE Trans. Green Commun. Netw.*, vol. 8, no. 1, pp. 375–390, 2024.

[5] M. Alawida, "A novel image encryption algorithm based on cyclic chaotic map in industrial IoT environments," *IEEE Trans. Ind. Informatics*, vol. 20, no. 8, pp. 10 530–10 541, 2024.

[6] Y. Ge, K. Xiong, Q. Wang, Q. Ni, P. Fan, and K. B. Letaief, "Aoi-minimal power adjustment in RF-EH-powered industrial IoT networks: A soft actor-critic-based method," *IEEE Trans. Mob. Comput.*, vol. 23, no. 9, pp. 8729–8741, 2024.

[7] C. Hu, J. Liu, H. Xia, S. Deng, and J. Yu, "A lightweight mutual privacy preserving K-means clustering in industrial IoT," *IEEE Trans. Netw. Sci. Eng.*, vol. 11, no. 2, pp. 2138–2152, 2024.

[8] J. Wang, Y. Hu, Y. Zhu, T. Wang, and A. Schmeink, "Reliability-optimal offloading for mobile edge-computing in low-latency industrial IoT networks," *IEEE Trans. Wirel. Commun.*, vol. 23, no. 10, pp. 12 765–12 781, 2024.

[9] Q. Xiong, X. Zhu, Y. Jiang, J. Cao, X. Xiong, and H. Wang, "Status prediction and data aggregation for AoI-oriented short-packet transmission in industrial IoT," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 611–625, 2023.

[10] L. Bai, X. Song, and L. Zhu, "Joint multi-feature information entity alignment for cross-lingual temporal knowledge graph with BERT," *IEEE Trans. Big Data*, vol. 11, no. 2, pp. 345–358, 2025.

[11] F. Yang, W. Chen, H. Lin, S. Wu, X. Li, Z. Li, and Y. Wang, "Task-oriented tool manipulation with robotic dexterous hands: A knowledge graph approach from fingers to functionality," *IEEE Trans. Cybern.*, vol. 55, no. 1, pp. 395–408, 2025.

[12] T. Zhang, J. Cheng, L. Miao, H. Chen, Q. Li, Q. He, J. Lyu, and L. Ma, "Multi-hop reasoning with relation based node quality evaluation for sparse medical knowledge graph," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 9, no. 2, pp. 1805–1816, 2025.

[13] Z. Liu, T. Sun, X. Sun, and W. Cui, "Estimating remaining useful life of aircraft engine system via a novel graph tensor fusion network based on knowledge of physical structure and thermodynamics," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–14, 2025.

[14] H. Sun, J. Wang, J. Weng, and W. Tan, "KG-ID: Knowledge graph-based intrusion detection on in-vehicle network," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 4, pp. 4988–5000, 2025.

[15] L. Liao, L. Zheng, J. Shang, X. Li, and F. Chen, "ATPF: An adaptive temporal perturbation framework for adversarial attacks on temporal knowledge graph," *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 3, pp. 1091–1104, 2025.

[16] T. Song, L. Yin, Y. Liu, L. Liao, J. Luo, and Z. Xu, "Expressiveness analysis and enhancing framework for geometric knowledge graph embedding models," *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 1, pp. 306–318, 2025.

[17] S. Bi, Z. Miao, and Q. Min, "LEMON: A knowledge-enhanced, type-constrained, and grammar-guided model for question generation over knowledge graphs," *IEEE Trans. Learn. Technol.*, vol. 18, pp. 256–272, 2025.

[18] C. Mai, Y. Chang, C. Chen, and Z. Zheng, "Enhanced scalable graph neural network via knowledge distillation," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 36, no. 1, pp. 1258–1271, 2025.

[19] X. Sun, B. Zhang, Y. Wang, J. Mai, Y. Wang, J. Tan, and W. Wang, "A multiscale attention mechanism super-resolution confocal microscopy for wafer defect detection," *IEEE Trans Autom. Sci. Eng.*, vol. 22, pp. 1016–1027, 2025.

[20] L. Jin and L. Guo, "Facial expression recognition based on feature representation learning and clustering-based attention mechanism," *IEEE Trans. Biom. Behav. Identity Sci.*, vol. 7, no. 2, pp. 182–194, 2025.

[21] Z. Wang, G. Wang, and G. He, "COSA plus: Enhanced co-operative systolic arrays for attention mechanism in transformers," *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*, vol. 44, no. 2, pp. 723–736, 2025.

[22] L. Zhang, J. Wang, Y. Guo, H. Zhang, L. Cheng, and W. Xia, "A self-attention mechanism-based model to detect IPv6 multi-field covert channels," *IEEE Trans. Cogn. Commun. Netw.*, vol. 11, no. 1, pp. 258–273, 2025.

[23] S. Li, L. Zhang, H. Guo, J. Li, J. Yu, X. He, Y. Zhao, and X. He, "CSA-FCN: Channel- and spatial-gated attention mechanism based fully complex-valued neural network for system matrix calibration in magnetic particle imaging," *IEEE Trans. Computational Imaging*, vol. 11, pp. 65–76, 2025.

[24] X. Liu, R. Dong, and H. Yang, "DHHG-TAC: Fusion of dynamic heterogeneous hypergraphs and transformer attention mechanism for visual question answering tasks," *IEEE Trans. Ind. Informatics*, vol. 21, no. 1, pp. 545–553, 2025.

[25] M. Li, Z. Wang, K. Li, and X. Liao, "Multiagent-system-based attention mechanism for predicting product popularity: Handling positive-negative diffusion over social networks," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 36, no. 1, pp. 1925–1932, 2025.

[26] Z. Fu, K. Jiang, Y. Xu, Y. Wang, T. Wen, H. Gao, Z. Zhong, and D. Yang, "Top-down attention-based mechanisms for interpretable autonomous driving," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 2, pp. 2212–2226, 2025.

[27] X. Li, W. Jiao, Q. Han, and Z. Lu, "Detection of FDIA in power grid based on hypergraph and attention mechanism," *IEEE Trans. Smart Grid*, vol. 16, no. 2, pp. 1862–1871, 2025.

[28] X. Xu, W. Lv, W. Wang, Y. Zhang, and J. Chen, "Empowering semantic segmentation with selective frequency enhancement and attention mechanism for tampering detection," *IEEE Trans. Artif. Intell.*, vol. 5, no. 6, pp. 3270–3283, 2024.