

A Similarity-Aware Graph Transformer-enhanced Probabilistic Case-based Reasoning Model for Knowledge Graph Reasoning

Yuejia Wu and Jian-tao Zhou*

College of Computer Science, Inner Mongolia University
Engineering Research Center of Ecological Big Data, Ministry of Education
Inner Mongolia Engineering Laboratory for Cloud Computing and Service Software
Inner Mongolia Engineering Laboratory for Big Data Analysis Technology
Hohhot, China
wuyuejia@imudges.com and cszjtao@imu.edu.cn

Abstract

Knowledge Graph Reasoning (KGR) is an effective way to ameliorate sparsity and incompleteness problems by inferring new knowledge based on existing knowledge. The probabilistic case-based reasoning (CBR) model can gather reasoning paths from similar entities and relations in KG, thus outperforming rule-based and embedding-based KGR methods. However, it is still limited by some problems, such as insufficient learning of similarity features and sparse intermediate activations. This paper proposes a Similarity-Aware graph transformer-enhanced probabilistic CBR model for **KGR**, namely SA-KGR. The proposed model regards the reasoning task as the KG query answering and is composed of two phases. The first phase is similarity-aware graph transformer-based graph feature encoding, which equips the similarity matrix and Mixture-of-Expert network to obtain fine-grained similarity features that are more helpful for reasoning path generation. The second phase is similarity-enhanced probabilistic case-based reasoning, which can retrieve and infer query answers from the generated candidate paths to complete brain-like cognitive reasoning. Extensive experimental results on various benchmarks unambiguously demonstrate that the proposed SA-KGR model can obtain the state-of-the-art results of current CBR-based methods.

Keywords: Knowledge Graph Reasoning; Case-based Reasoning; Graph Transformer; Mixture-of-Expert network

Introduction

The Knowledge Graph (KG) is essentially a semantic network that reveals the relationships between entities, which represents and stores knowledge in the form of (*headentity, relation, tailentity*) facts. Existing KGs such as WordNet (Miller, 1995), NELL (Carlson et al., 2010), and Freebase (Bollacker, Evans, Paritosh, Sturge, & Taylor, 2008) are usually sparse and incomplete, limiting their performance in Recommendation System (J.-C. Zhang, Zain, Zhou, Chen, & Zhang, 2024), Knowledge-Based Question Answering (Z. Zhang, Wen, & Zhao, 2025), Anomaly Detection (Mitropoulou, Kokkinos, Soumplis, & Varvarigos, 2024), etc.

Knowledge Graph Reasoning (KGR) is an effective approach to improve those problems by predicting new knowledge according to the existing knowledge in the KG. There are four main objectives of KGR: link prediction (Cao & Luo, 2025), fact prediction (Padia, Kalpakis, Ferraro, & Finin, 2019), entity prediction (Zou, An, & Li, 2022), and relation prediction (Cui, Kapanipathi, Talamadupula, Gao, & Ji,

2021). Among them, the link prediction task uses a scoring function to calculate the probability of the given relation belonging to the KG, which can significantly improve KG's quality, but its accuracy needs to be further improved. Therefore, this paper will focus on improving the performance of the link prediction task. Recent studies have shown that the case-based reasoning (CBR) methods (Dubitzky, Büchner, & Azuaje, 1999; Bartlmae & Riemenschneider, 2000; Das, Godbole, Dhuliawala, Zaheer, & McCallum, 2020; Das, Godbole, Monath, Zaheer, & McCallum, 2020; Wu & Zhou, 2023) can mimic the cognitive processes of the human brain, i.e., solving a new problem by retrieving “cases” that are similar to the existing problem, so that can achieve better results more than rule-based and embedding-based reasoning methods. However, there are still existing problems such as insufficient learning of similarity features and sparse intermediate activations.

In this paper, we propose a KGR model based on a similarity-aware graph transformer-enhanced probabilistic CBR model, referred to as SA-KGR. Our approach regards the link prediction task as the KG query answering and consists of two phases. The first phase is similarity-aware graph transformer-based graph feature encoding (SAFE), which can obtain finer-grained feature information while improving the model's generalization. The second phase is similarity-enhanced probabilistic case-based reasoning (SE-CBR), which can further mimic the cognitive reasoning process of the human brain. The main contributions of this paper are as follows:

- We propose a two-phase similarity-aware graph transformer-enhanced probabilistic case-based reasoning model for KGR. To the best of our current knowledge, this is the first approach that designs a global similarity matrix into the attention calculation process in the field of CBR-based KGR, which can obtain finer-grained feature information and generate more effective candidate reasoning paths to complete the reasoning.
- We present and equip a Mixture-of-Expert (MoE) network in the similarity-aware graph transformer-based graph feature encoding method proposed in this paper, which can improve the model's generalization and scalability ability and make it handle exponentially increasing parameters to maximize computational efficiency.

*Corresponding author.

- We conduct extensive contrast experiments on three standard KGs and the results illustrate that the proposed KGR model SA-KGR can effectively enhance the performance of subsistent CBR-based models and obtain the state-of-the-art experimental results.

Related Work

This section first briefly explains the sparsity and incompleteness problems in the KG and then states the KGR solutions based on the CBR approach. As shown in Fig. 1, we give the ubiquitous description of sparsity and incompleteness in KG: (i) KG’s sparsity: usually includes entity and relation sparsity, meaning that the total number of facts is far more than the number of entities and relations. For example, in the KG shown in Fig. 1, it can be seen that the ratio of entities’ kind (7) and relations’ kind (3) to the number of facts (10) is small. (ii) KG’s incompleteness: represents missing the whole fact or missing one part of it. As shown in Fig. 1, many entities and relations are missed in the KG, such as the people related to “Lionel Messi” and others, the “brother” relationship of “Mateo Messi” and “Thiago Messi”, etc.

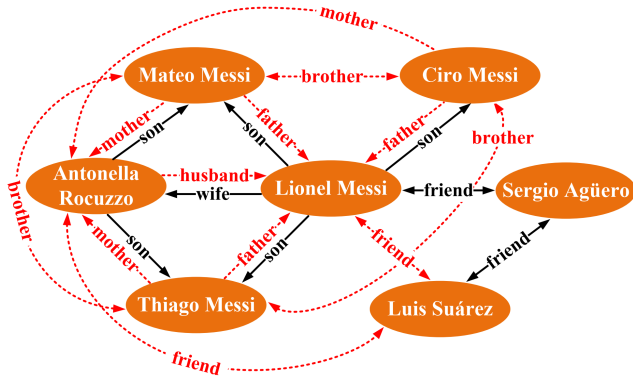


Figure 1: An example of KG. The solid black lines denote existing relations, and the dashed red lines represent missing relations.

KGR is an effective way to improve those problems. Recent studies have demonstrated that CBR-based approaches can achieve more advanced prediction results compared to other rule-based and embedding-based models. Das et al. (Das, Godbole, Dhuliawala, et al., 2020; Das, Godbole, Monath, et al., 2020) was the first to frame the reasoning task as a KG query answering, i.e. answering questions in the form of (“Mateo Messi”, “father”, ?). Then, retrieved k similar entities (cases) to the query entity and generated multiple KG paths, which served as solutions connecting the retrieved cases through the query relation. However, those methods still face challenges, such as insufficient graph feature extraction and omission of contextual relation information. To address these limitations, Wu et al. (Wu & Zhou, 2023) proposed CICBR, a contextual information-augmented probabilistic CBR model for KGR. This model provides a more

fine-grained representation of entity and relation features and is the first to introduce a Graph Transformer for KG representation and learning, achieving state-of-the-art (SOTA) results among current CBR-based methods.

However, current CBR-based models face two critical limitations. First, insufficient learning of similarity features (e.g., nuanced semantic differences in relationships such as “father” versus “stepfather”) during KG feature acquisition, which results in noisy reasoning paths, e.g., irrelevant entity connections that degrade prediction accuracy. Second, sparse intermediate activations in Feed-Forward Network (FFN) layers. While increasing embedding dimensions in the encoding phase escalates model parameters (due to the quadratic scaling of attention mechanisms), most intermediate neuron activations remain near zero. This sparsity undermines training efficiency by limiting gradient flow and discarding critical latent features, ultimately hindering the model’s ability to generalize across complex relational patterns. This paper focuses on those problems and proposes a KGR model based on a similarity-aware graph transformer-enhanced probabilistic CBR model named SA-KGR that can capture global fine-grained similarity feature information on the basis of improving the scalability and generalization ability of the model, so as to improve the accuracy of obtaining the correct query answer.

Methodology

Preliminaries

To facilitate the understanding of the subsequent formulae, the general definition of KG is defined as follows:

Definition 1 (Knowledge Graph). A KG $\mathcal{G} = (V, E, \mathcal{R})$ is a collection of facts stored as triplets (e_1, r, e_2) , where V is the set of entities, \mathcal{R} is the set of binary relations, $E \subseteq V \times \mathcal{R} \times V$ is the edges of the KG, and $e_1, e_2 \in V, r \in \mathcal{R}$. Following previous settings(?, ?), for a fact $(e_1, r, e_2) \in E$, we add its reverse fact (e_2, r^{-1}, e_1) to the KG whatever the set of binary relations \mathcal{R} contains the inverse relation r^{-1} .

Then, we regard the KGR as a query answering task on KG, i.e., for a question $(e_{1q}, r_q, ?)$, the model needs to give the corresponding answer, which is an entity in the KG.

Definition 2 (KG Query Answering Task). For an input query $(e_{1q}, r_q, ?)$, starting from the entity corresponding to e_{1q} in \mathcal{G} , the KG query answering model stops at a entity that is predicts as the answer following a path in the graph.

Following previous settings(Das, Godbole, Monath, et al., 2020), the paths used in **Definition 2** are defined as follows:

Definition 3 (KG Paths). A KG path $p = (e_1, r_1, e_2, r_2, \dots, r_n, e_{n+1})$ between two entities e_1 and e_{n+1} is defined as a sequence of alternating entities and relations that connect them, where $st(p) = e_1$, $en(p) = e_{n+1}$, $len(p) = n$, and $type(p) = (r_1, r_2, \dots, r_n)$. Let \mathcal{P} is the set of all paths in \mathcal{G} and $\mathcal{P}_n \subseteq \mathcal{P} = \{p \mid len(p) \leq n\}$ is the set of all paths of length up to n . Let $P_n = \{type(p) \mid p \in \mathcal{P}_n\}$ represents the set of all path types with length up to n and $P_n(e_1, r) \subseteq P_n$ is all path types of length up to n that originate at e_1 by a

direct edge of type r , i.e., if $S_{e_1 r} = \{e_2 \mid (e_1, r, e_2) \in \mathcal{G}\}$ is the set of entities that are connected to e_1 via a direct edge r , then $P_n(e_1, r)$ is the set of all path types of length up to n that start from e_1 and end at entities in $S_{e_1 r}$. Besides, $\mathcal{P}_n(e_1, r)$ is defined to denote paths instead of path types.

The proposed model SA-KGR

This section begins with an overview of the model’s architecture, followed by detailed discussions of its core components in the subsequent two subsections.

Overview

To improve the problems mentioned above, we propose the SA-KGR model to approach KGR. As shown in Fig. 2, the proposed SA-KGR consists of two phases: (a) similarity-aware graph transformer-based graph feature encoding (SAFE) and (b) similarity-enhanced probabilistic case-based reasoning (SE-CBR). Specifically, given an input KG and a query, SA-KGR’s first phase SAFE is to encode the graph feature representations with fine-grained similarity features and then design an MoE network to use the parameter sparsity for improving the model’s generalization. Next, SA-KGR’s second phase SE-CBR is to generate reasoning KG paths from the entities and relations that are similar to the query entity and relation respectively, and then calculate the score of each answer candidate and weigh paths with an estimate of their frequency and precision.

The Similarity-Aware Graph Transformer-based Graph Feature Encoding

As the first step in model processing, it is importance to design a graph feature encoding method that can effectively characterize the KG. However, the encoding methods in the previous CBR-based KGR model are relatively simple or utilized the Graph Transformer model (Dwivedi & Bresson, 2020) to calculate attention scores, which leads to the problem of insufficient learning of similar features. To improve this problem, we propose the similarity-aware graph transformer-based graph feature encoding method named SAFE into the CBR field to capture more fine-grained similarity feature information in the KG.

Specifically, following previous approaches (Dwivedi & Bresson, 2020), we pre-compute the Laplacian eigenvectors of KGs, which are defined as follows:

$$\Delta = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = U^T \Lambda U, \quad (1)$$

where A denotes a $|V| \times |V|$ adjacency matrix, D represents a degree matrix, and Λ, U correspond to the eigenvalues and eigenvectors respectively.

After pre-computing the Laplacian eigenvectors, for a Graph \mathcal{G} with entity features $\alpha_{e_i} \in \mathbb{R}^{d_n \times 1}$ for each entity e_i , the input features α_{e_i} is passed through a linear projection to embed it to d -dimensional hidden features $h_{e_i}^0$.

$$\hat{h}_{e_i}^0 = A^0 \alpha_{e_i} + a^0, \quad (2)$$

where $A^0 \in \mathbb{R}^{d \times k}$ and $a^0 \in \mathbb{R}^d$ are the parameters of the linear

projection layers. Then, we embed the pre-computed positional encoding of dim k via a linear projection and add it to the features $\hat{h}_{e_i}^0$.

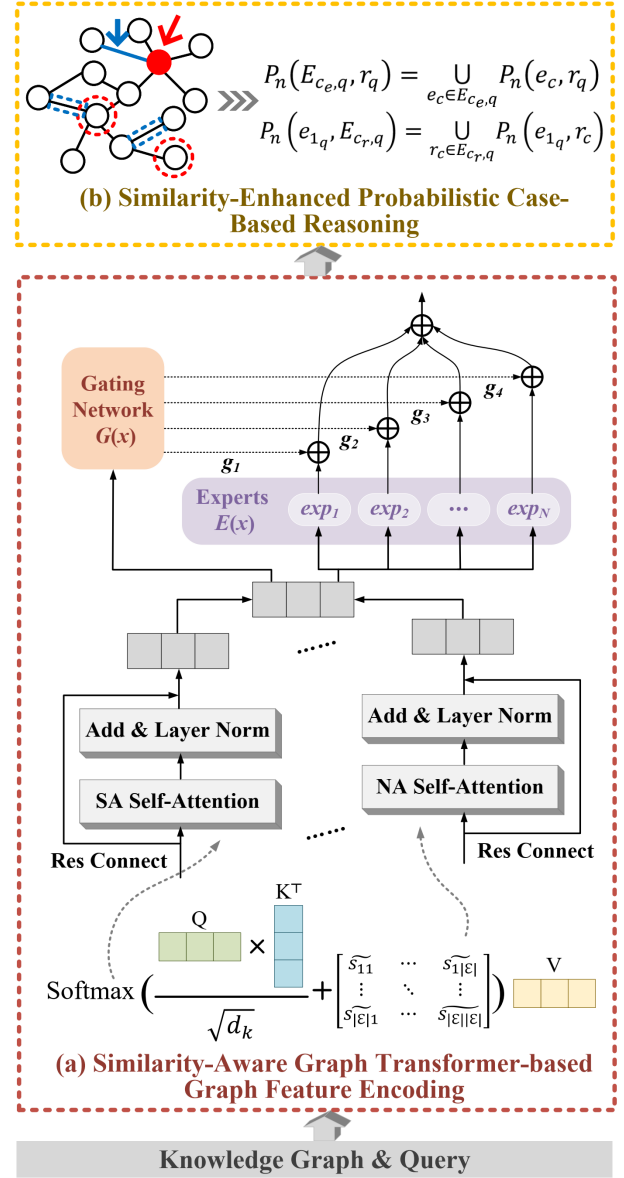


Figure 2: Overview of the proposed SA-KGR model.

$$h_{e_i}^0 = \hat{h}_{e_i}^0 + C^0 \lambda_{e_i} + c^0, \quad (3)$$

where $C^0 \in \mathbb{R}^{d \times k}$ and $c^0 \in \mathbb{R}^d$.

To simultaneously capture semantic relevance in attention mechanisms and quantify entity importance, we add the entity in-degree and out-degree characteristics into the $\hat{h}_{e_i}^0$ and refined it as the \mathbf{h}'_i . The specific computations are formalized as follows:

$$\mathbf{h}'_i = \hat{h}_{e_i}^0 + \mathbf{z}^-_{deg^-(i)} + \mathbf{z}^+_{deg^+(i)}, \quad (4)$$

where $\mathbf{z}^-, \mathbf{z}^+ \in \mathbb{R}^d$ are learnable embedding vectors specified

by the in-degree $deg^-(i)$ and out-degree $deg^+(i)$ respectively, and d is hidden dimension.

Then, we use the multi-head self-attention mechanism Transformer (Vaswani, 2017) to capture more feature information, the entity representations $\mathbf{H} = \{\mathbf{h}'_1, \mathbf{h}'_2, \dots, \mathbf{h}'_{|V|}\} \in \mathbb{R}^{|V| \times d}$ is projected onto corresponding representations \mathbf{Q} , \mathbf{K} and \mathbf{V} by three matrices $\mathbf{W}_Q \in \mathbb{R}^{d \times d_Q}$, $\mathbf{W}_K \in \mathbb{R}^{d \times d_K}$, $\mathbf{W}_V \in \mathbb{R}^{d \times d_V}$. The self-attention is calculated as follows:

$$\mathbf{Q} = \mathbf{H}\mathbf{W}_Q, \mathbf{K} = \mathbf{H}\mathbf{W}_K, \mathbf{V} = \mathbf{H}\mathbf{W}_V. \quad (5)$$

However, there still exists the problem of insufficient learning of similarity features. To improve this problem, we design a global similarity matrix and add it into the attention calculation process as follows:

$$\text{Attn}(\mathbf{H}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_K}} + \tilde{S}\right)\mathbf{V}, \quad (6)$$

where $\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_K}}$ calculates the similarity between query \mathbf{Q} and key \mathbf{K} , and $\tilde{S} \in \mathbb{R}^{|V| \times |V|}$ represents a global similarity-aware matrix that indicates the importance and influence between entities in the KG. The calculation of S is shown as follows:

$$S = \begin{bmatrix} S_{11} & \dots & S_{1|V|} \\ \vdots & \ddots & \vdots \\ S_{|V|1} & \dots & S_{|V||V|} \end{bmatrix}, \quad (7)$$

where $S_{ij} \in S$ is an entity similarity score between entities i and j , which is obtained as follows:

$$S_{ij} = \frac{h_i \cdot h_j}{|h_i| |h_j|}. \quad (8)$$

To make S_{ij} learnable and to obtain more inferential similarity information, we enhance and substitute it as follows:

$$\tilde{S}_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{W}_1 S_{ij}))}{\sum_{k \in \mathcal{N}_i} \exp(\text{LeakyReLU}(\mathbf{W}_2 S_{ik}))}, \quad (9)$$

where $\mathbf{W}_1, \mathbf{W}_2 \in \mathbb{R}^{|V| \times |V|}$ are learnable parameters.

After obtaining the one-head self-attention, we concat multi-head self-attention calculation through $\mathbf{W}^O \in \mathbb{R}^{d \times d}$, which can be expressed as follows:

$$\text{MultiAttn}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)\mathbf{W}^O, \quad (10)$$

$$\text{head}_i = \text{Attn}(\mathbf{Q}\mathbf{W}_i^Q, \mathbf{K}\mathbf{W}_i^K, \mathbf{V}\mathbf{W}_i^V),$$

where h is the number of scaled dot-product attention.

Finally, after obtaining all the attention header information, we output it to a fully connected feed-forward network (FFN). However, due to the need to jointly calculate a large number of entity similarity information, the model parameters surge while enhancing the learning ability of the model, but most of the elements in it are zero values. Therefore, inspired by Shazeer et al. (Shazeer et al., 2017), we design an MoE network for the multi-head attention mechanism. Specifically,

each MoE network consists of a set of N expert networks E and a gate network G , which can be formulated as follows:

$$y = \sum_{i=1}^N E_i(x)G(x)_i = \sum_{i=1}^N \text{FFN}_i(x)G(x)_i \quad (11)$$

$$= \sum_{i=1}^N (\max(0, x\mathbf{W}_1 + \mathbf{b}_1)\mathbf{W}_2 + \mathbf{b}_2)_i G(x)_i.$$

We multiply Top-K experts with most weight by a learnable weight matrix $\mathbf{W}_g \in \mathbb{R}^{d \times N}$ and utilize a *Softmax* function as the final output, which is calculated as follows:

$$G(x) = \text{Softmax}(\text{Top} - K(x \cdot \mathbf{W}_g)), \text{ where } K < N. \quad (12)$$

The Similarity-Enhanced Probabilistic Case-Based Reasoning

After obtaining KG features representation, we first use the Cosine similarity calculation formula to find similar entities to the query entity that has at least a relation r_q , and similar relations linked to the query entity e_{1q} . Therefore, we let $E_{c_e, q}$ and $E_{c_r, q}$ denote the set of contextual entities and contextual relations for the query q respectively.

Then, we give the representation of the entity retrieved corresponding to the contextual information as follows:

$$P_n(E_{c_e, q}, r_q) = \bigcup_{e_c \in E_{c_e, q}} P_n(e_c, r_q) \quad (13)$$

$$P_n(e_{1q}, E_{c_r, q}) = \bigcup_{r_c \in E_{c_r, q}} P_n(e_{1q}, r_c) \quad (14)$$

Then, for a given query, the probability of finding the answer entity e_2 is defined as follows:

$$P(e_2 | e_{1q}, r_q) = \sum_{p \in P_n(E_{c_e, q}, r_q) \cup P_n(e_{1q}, E_{c_r, q})} P(e_2, p | e_{1q}, r_q)$$

$$= \sum_p P(p | e_{1q}, r_q) P(e_2 | p, e_{1q}, r_q) \quad (15)$$

Next, we set the random variable representing the path types obtained from $E_{c_e, q}$ and $E_{c_r, q}$. $P(p | e_{1q}, r_q)$ represents the probability of finding a path type, which denotes how often each path type appears with the query and represents the prior probability of the path type. $P(e_2 | p, e_{1q}, r_q)$ is the proportion of times, which can capture the likelihood of reaching the right answer or the ‘‘precision’’ of a path type.

Similar with the previous approaches, c is a random variable that represents the cluster assignment of the query entity. Then for the path-prior term, there have

$$P(p | e_{1q}, r_q) = \sum_c P(c | e_{1q}, r_q) P(p | c, e_{1q}, r_q) \quad (16)$$

where $P(c | e_{1q}, r_q)$ is zero for all clusters except the cluster to which the query entity belongs, and $P(p | c_{e_{1q}}, e_{1q}, r_q) = P(p | c_{e_{1q}}, r_q)$ if $c_{e_{1q}}$ is the cluster in which the e_{1q} has been

assigned. Then, the path prior $P(p \mid c, r_q)$ is calculated as follows:

$$\frac{\sum_{e_c \in c} \sum_{p' \in \mathcal{P}_n(e_c, r_q) \cup \mathcal{P}_n(e_{1_q}, r_c)} \mathbb{1}[\text{type}(p') = p]}{\sum_{e_c \in c} \sum_{p' \in \mathcal{P}_n(e_c, r_q) \cup \mathcal{P}_n(e_{1_q}, r_c)} \mathbb{1}} \quad (17)$$

For each entity within cluster c , we analyze paths connecting e_c to its directly associated entities through relation type r_q and its contextual relations E_{c_r} . The path prior for a specific path type p is computed as the frequency ratio of the path type p occurring in $\mathcal{P}_n(e_c, r_q)$, and the path-precision probability ($P(e_2 \mid p, c, r_q)$) is defined as follows:

$$\frac{\sum_{e_c \in c} \sum_{p' \in \mathcal{P}_n(e_c)} \mathbb{1}[\text{type}(p') = p] \cdot \mathbb{1}[en(p') \in S_{e_c r_q}]}{\sum_{e_c \in c} \sum_{p' \in \mathcal{P}_n(e_c)} \mathbb{1}[\text{type}(p') = p]} \quad (18)$$

where $\mathcal{P}_n(e_c)$ is the paths starting from the entity e_c up to length n , $en(p)$ is the end entity for a path p and $S_{e_c r_q}$ is the set of entities connecting to e_c via a direct relation of type r_q . Then, given r_q , Eq.18 quantifies the success rate of path type p in reaching valid answer entities when traversing from the cluster entity e_c .

Experiments and Results

This section conducts extensive comparison experiments on three standard KGs to verify the performance of the proposed SA-KGR model.

Experimental Datasets

To verify the effectiveness of the proposed CBR-based KGR model SA-KGR, we utilize three different KGR standard datasets: NELL-995 (Xiong, Hoang, & Wang, 2017), FB122 (Guo, Wang, Wang, Wang, & Guo, 2016), and WN18RR (Dettmers, Minervini, Stenetorp, & Riedel, 2018), which as shown in Table 1.

Table 1: Overview of the experimental datasets.

| Dataset | #Ent | #Rel | #Fact |
|----------|--------|------|---------|
| NELL-995 | 75,492 | 200 | 154,213 |
| FB122 | 9,738 | 122 | 112,476 |
| WN18RR | 40,943 | 11 | 93,003 |

Baselines and Evaluation Metrics

The experiment represents the rule-based, embedding-based, and CBR-based methods as baselines comparing our proposed model SA-KGR to prove its validity thoroughly. The rule-based baselines consist of KALE (Guo et al., 2016), ASR (Minervini, Demeester, Rocktäschel, & Riedel, 2017), and KG_{LR} (García-Durán & Niepert, 2017). The embedding-based baselines contain TransE (Bordes, Usunier, Garcia-Duran, Weston, & Yakhnenko, 2013), DistMult (Yang, Yih, He, Gao, & Deng, 2015), ComplEx (Trouillon, Welbl, Riedel, Gaussier, & Bouchard, 2016), ConvE (Dettmers et al., 2018), RotatE (Sun, Deng, Nie, & Tang, n.d.), GNTP (Minervini, Bošnjak, Rocktäschel, Riedel, & Grefenstette, 2020), and

Table 2: Overall results of link prediction task on FB122 datasets (Test-I and Test-II).

| Models | Test-I | | | | Test-II | | | |
|---------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | Hits@K | | | | Hits@K | | | |
| | 3 | 5 | 10 | MRR | 3 | 5 | 10 | MRR |
| KALE-Pre [†] | 0.358 | 0.419 | 0.498 | 0.291 | 0.829 | 0.861 | 0.899 | 0.713 |
| KALE-Joint [†] | 0.384 | 0.447 | 0.522 | 0.325 | 0.797 | 0.841 | 0.896 | 0.684 |
| ASR-DistMult [†] | 0.363 | 0.403 | 0.449 | 0.330 | 0.980 | 0.990 | 0.992 | 0.948 |
| ASR-ComplEx [†] | 0.373 | 0.410 | 0.459 | 0.338 | 0.992 | 0.993 | 0.994 | 0.984 |
| TransE [†] | 0.360 | 0.415 | 0.481 | 0.296 | 0.775 | 0.828 | 0.884 | 0.630 |
| DistMult [†] | 0.360 | 0.403 | 0.453 | 0.313 | 0.923 | 0.938 | 0.947 | 0.874 |
| ComplEx [†] | 0.370 | 0.413 | 0.462 | 0.329 | 0.914 | 0.919 | 0.924 | 0.887 |
| RotatE [†] | 0.511 | 0.551 | 0.603 | 0.471 | 0.868 | 0.886 | 0.907 | 0.846 |
| GNTPs [†] | 0.337 | 0.369 | 0.412 | 0.313 | 0.982 | 0.990 | 0.993 | 0.977 |
| CBR [†] | 0.400 | 0.445 | 0.488 | 0.359 | 0.678 | 0.718 | 0.759 | 0.636 |
| PCBR [†] | 0.490 | 0.527 | 0.571 | 0.457 | 0.948 | 0.950 | 0.953 | 0.948 |
| CICBR | 0.508 | 0.543 | 0.592 | 0.465 | 0.959 | 0.971 | 0.975 | 0.956 |
| SA-KGR(Ours) | 0.513 | 0.545 | 0.597 | 0.466 | 0.962 | 0.975 | 0.977 | 0.960 |

MINERVA (Das et al., 2018). The CBR-based baselines include CBR (Das, Godbole, Dhuliawala, et al., 2020), PCBR (Das, Godbole, Monath, et al., 2020), and CICBR (Wu & Zhou, 2023). Because the research of CBR-based KGR is still in the development stage, there are few baselines for comparison.

In the link prediction task, two kinds of standard metrics were used to evaluate the experimental performance: Mean Reciprocal Ranking (MRR) and Hits@K. For each metric, a higher score indicates a better effect. The MRR is defined as follows:

$$MRR = \frac{1}{|N|} \sum_{i=1}^{|N|} \frac{1}{rank_i} = \frac{1}{|N|} \left(\frac{1}{rank_1} + \frac{1}{rank_2} + \dots + \frac{1}{rank_{|N|}} \right) \quad (19)$$

N is the set of triples and $rank_i$ is the link prediction ranking of the i -th triple.

In addition, $Hits@K$ ($K = 1, 3, 5, 10$) is defined as follows:

$$Hits@K = \frac{1}{|N|} \sum_{i=1}^{|N|} \mathbb{I}(rank_i \leq K) \quad (20)$$

where $\mathbb{I}(\cdot)$ is an indicator function that the value sets to 1 if the condition is true, otherwise it sets to 0.

Results and Analysis

The link prediction overall results for the three KG datasets are shown in Table 2, Table 3, and Table 4, where “[†]” means the results taken from literature (Das, Godbole, Monath, et al., 2020). It can be found that the proposed SA-KGR model is obviously improved compared with CBR-based KGR methods that the SA-KGR model can improve the {Hits@3, Hits@5, Hits@10, MRR} average reasoning accuracy compared with CBR-based baselines by {12.315%, 10.469%, 9.139%, 14.094%} under the FB122 dataset and improve the {Hits@1, Hits@3, Hits@10, MRR} average reasoning accuracy compared with CBR-based baselines under the NELL-995 and WN18RR datasets by {8.712%, 4.357%, 3.031%, 8.819%} and {16.474%, 7.722%, 14.952%, 16.778%} respectively.

Table 3: Overall results of link prediction task on FB122 datasets (Test-ALL).

| Models | Test-ALL | | | MRR |
|-------------------------------|--------------|--------------|--------------|--------------|
| | Hits@3 | Hits@5 | Hits@10 | |
| KALE-Pre [†] | 0.617 | 0.662 | 0.718 | 0.523 |
| KALE-Joint [†] | 0.612 | 0.664 | 0.728 | 0.523 |
| ASR-DistMult [†] | 0.707 | 0.731 | 0.752 | 0.675 |
| ASR-CompLex [†] | 0.717 | 0.736 | 0.757 | 0.698 |
| KG _{LR} [†] | 0.740 | 0.770 | 0.797 | 0.702 |
| TransE [†] | 0.589 | 0.642 | 0.702 | 0.480 |
| DistMult [†] | 0.674 | 0.702 | 0.729 | 0.628 |
| CompLex [†] | 0.673 | 0.695 | 0.719 | 0.641 |
| RotatE [†] | 0.708 | 0.736 | 0.770 | 0.678 |
| GNTPs [†] | 0.692 | 0.711 | 0.732 | 0.678 |
| CBR [†] | 0.570 | 0.612 | 0.653 | 0.527 |
| PCBR [†] | 0.742 | 0.760 | 0.782 | 0.727 |
| CICBR | 0.749 | 0.771 | 0.788 | 0.733 |
| SA-KGR (Ours) | 0.751 | 0.775 | 0.792 | 0.738 |

Table 4: Overall results of link prediction task on NELL-995 and WN18RR datasets.

| Models | NELL-995 | | | | WN18RR | | | |
|-----------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | Hits@K | | | | Hits@K | | | |
| | 1 | 3 | 10 | MRR | 1 | 3 | 10 | MRR |
| TransE [†] | 0.53 | 0.79 | 0.87 | 0.67 | - | - | 0.50 | 0.23 |
| DistMult [†] | 0.61 | 0.73 | 0.79 | 0.68 | 0.39 | 0.44 | 0.49 | 0.43 |
| CompLex [†] | 0.61 | 0.76 | 0.83 | 0.69 | 0.41 | 0.46 | 0.51 | 0.44 |
| ConvE [†] | 0.67 | 0.81 | 0.86 | 0.75 | 0.40 | 0.44 | 0.52 | 0.43 |
| RotatE [†] | 0.65 | 0.82 | 0.87 | 0.74 | 0.43 | 0.49 | 0.57 | 0.48 |
| GNTP [†] | - | - | - | - | 0.41 | 0.44 | 0.48 | 0.43 |
| MINERVA [†] | 0.66 | 0.77 | 0.83 | 0.72 | 0.40 | 0.43 | 0.49 | 0.43 |
| CBR [†] | 0.70 | 0.83 | 0.87 | 0.77 | 0.38 | 0.46 | 0.51 | 0.43 |
| PCBR [†] | 0.77 | 0.85 | 0.89 | 0.81 | 0.43 | 0.49 | 0.55 | 0.48 |
| CICBR | 0.80 | 0.88 | 0.89 | 0.85 | 0.46 | 0.50 | 0.59 | 0.51 |
| SA-KGR (Ours) | 0.82 | 0.89 | 0.91 | 0.88 | 0.49 | 0.52 | 0.63 | 0.55 |

Compared with CBR-based baselines, the link prediction results on all experimental KG datasets of the SA-KGR model proposed in this paper are explicitly improved. Besides, under the FB122 (Test-ALL), NELL-995, and WN18RR, our proposed SA-KGR model can achieve the SOTA performance against the rule-based and embedding-based methods. That is because: (i) by designing a global similarity matrix into the attention calculation process, the model can obtain more finer-grained feature information and generate more effective candidate reasoning paths to complete the reasoning; (ii) by presenting and equipping an MoE network in the proposed feature encoding method, the model’s generalization and scalability ability can be improved that can handle exponentially increasing parameters to maximize computational efficiency, so as to improve the accuracy of model’s reasoning precision performance.

In general, comparing and analyzing the above experimental results, it is demonstrated that more plentiful fine-grained graph features can be achieved by using the proposed SA-KGR model, to generate a path more conducive to inference. Moreover, the sparsity of intermediate activation values can be utilized to improve the parameter capacity of the model, thus effectively improving the inference accuracy of

the model and obtaining SOTA results on all experimental KG datasets.

Conclusion

In this paper, we proposed a similarity-aware graph transformer-enhanced probabilistic case-based reasoning model for KGR named SA-KGR. The SA-KGR first designed a similarity-aware graph transformer-based graph feature encoding method, which can effectively improve the generalization and scalability of the model to obtain more fine-grained global similarity information by adding a global similarity matrix into the multi-head self-attention calculation process and designing the MoE network for the FFN layer. Then, the SA-KGR presented a similarity-enhanced probabilistic case-based reasoning method, which can imitate the cognitive process of the human brain to obtain the query answer by calculating the probability from candidate KG paths. Finally, extensive comparison experiments on three standard KGs illustrated that our proposed SA-KGR model could achieve the SOTA performance compared with other existing CBR-based baselines.

Although our proposed model can improve current CBR-based KGR methods, how to obtain more effective feature information needs further study. In the future, we will design CBR methods that are more in line with the human brain to improve the model’s reasoning ability.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant No.62472236 and No.62162046, the Inner Mongolia Natural Science Foundation under Grant No.2019GG372, the Inner Mongolia Science and Technology Project under Grant No.2021GG0155, the Natural Science Foundation of Major Research Plan of Inner Mongolia under Grant No.2019ZD15, the Inner Mongolia University Youth Academic Talent Research Launch Project under Grant No.10000-23112101/247, and the fund of Supporting the Reform and Development of Local Universities (Disciplinary Construction) and the special research project of First-class Discipline of Inner Mongolia A. R. of China under Grant YLXKZX-ND-036.

References

- Bartlmae, K., & Riemenschneider, M. (2000). Case based reasoning for knowledge management in kdd projects. In *Pakm*.
- Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008). Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 acm sigmod international conference on management of data* (pp. 1247–1250).
- 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.

- Cao, Z., & Luo, C. (2025). Link prediction for knowledge graphs based on extended relational graph attention networks. *Expert Systems with Applications*, 259, 125260.
- Carlson, A., Betteridge, J., Kisiel, B., Settles, B., Hruschka, E., & Mitchell, T. (2010). Toward an architecture for never-ending language learning. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 24, pp. 1306–1313).
- Cui, Z., Kapanipathi, P., Talamadupula, K., Gao, T., & Ji, Q. (2021). Type-augmented relation prediction in knowledge graphs. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 35, pp. 7151–7159).
- Das, R., Dhuliawala, S., Zaheer, M., Vilnis, L., Durugkar, I., Krishnamurthy, A., . . . McCallum, A. (2018). Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning. In *International conference on learning representations*.
- Das, R., Godbole, A., Dhuliawala, S., Zaheer, M., & McCallum, A. (2020). A simple approach to case-based reasoning in knowledge bases. *arXiv preprint arXiv:2006.14198*.
- Das, R., Godbole, A., Monath, N., Zaheer, M., & McCallum, A. (2020). Probabilistic case-based reasoning for open-world knowledge graph completion. In *Findings of the association for computational linguistics: Emnlp 2020* (pp. 4752–4765).
- Dettmers, T., Minervini, P., Stenetorp, P., & Riedel, S. (2018). Convolutional 2d knowledge graph embeddings. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 32).
- Dubitzky, W., Büchner, A., & Azuaje, F. (1999). Viewing knowledge management as a case-based reasoning application. In *Aaai workshop technical report ws-99-10* (pp. 23–27).
- Dwivedi, V. P., & Bresson, X. (2020). A generalization of transformer networks to graphs. *arXiv preprint arXiv:2012.09699*.
- García-Durán, A., & Niepert, M. (2017). Kblrn: End-to-end learning of knowledge base representations with latent, relational, and numerical features. *arXiv preprint arXiv:1709.04676*.
- Guo, S., Wang, Q., Wang, L., Wang, B., & Guo, L. (2016). Jointly embedding knowledge graphs and logical rules. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 192–202).
- Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, 38(11), 39–41.
- Minervini, P., Bošnjak, M., Rocktäschel, T., Riedel, S., & Grefenstette, E. (2020). Differentiable reasoning on large knowledge bases and natural language. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 34, pp. 5182–5190).
- Minervini, P., Demeester, T., Rocktäschel, T., & Riedel, S. (2017). Adversarial sets for regularising neural link predictors. In *Uncertainty in artificial intelligence-proceedings of the 33rd conference, uai 2017*.
- Mitropoulou, K., Kokkinos, P., Soumplis, P., & Varvarigos, E. (2024). Anomaly detection in cloud computing using knowledge graph embedding and machine learning mechanisms. *Journal of Grid Computing*, 22(1), 6.
- Padia, A., Kalpakis, K., Ferraro, F., & Finin, T. (2019). Knowledge graph fact prediction via knowledge-enriched tensor factorization. *Journal of Web Semantics*, 59, 100497.
- Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., & Dean, J. (2017). Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*.
- Sun, Z., Deng, Z.-H., Nie, J.-Y., & Tang, J. (n.d.). Rotate: Knowledge graph embedding by relational rotation in complex space. In *International conference on learning representations*.
- Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., & Bouchard, G. (2016). Complex embeddings for simple link prediction. In *International conference on machine learning* (pp. 2071–2080).
- Vaswani, A. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*.
- Wu, Y., & Zhou, J.-t. (2023). A contextual information-augmented probabilistic case-based reasoning model for knowledge graph reasoning. In *International conference on case-based reasoning* (pp. 102–117).
- Xiong, W., Hoang, T., & Wang, W. Y. (2017). Deeppath: A reinforcement learning method for knowledge graph reasoning. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 564–573).
- Yang, B., Yih, S. W.-t., He, X., Gao, J., & Deng, L. (2015). Embedding entities and relations for learning and inference in knowledge bases. In *Proceedings of the international conference on learning representations (iclr) 2015*.
- Zhang, J.-C., Zain, A. M., Zhou, K.-Q., Chen, X., & Zhang, R.-M. (2024). A review of recommender systems based on knowledge graph embedding. *Expert Systems with Applications*, 123876.
- Zhang, Z., Wen, L., & Zhao, W. (2025). Rule-kbqa: Rule-guided reasoning for complex knowledge base question answering with large language models. In *Proceedings of the 31st international conference on computational linguistics* (pp. 8399–8417).
- Zou, C., An, J., & Li, G. (2022). Knowledge graph entity type prediction with relational aggregation graph attention network. In *European semantic web conference* (pp. 39–55).