

# Commonsense-Guided Inductive Relation Prediction with Dual Attention Mechanism

Yuxiao Duan , Jiuyang Tang <sup>\*</sup>, Hao Xu, Changsen Liu and Weixin Zeng 

Laboratory for Big Data and Decision, National University of Defense Technology, Changsha 410072, China; duanyuxiao19@nudt.edu.cn (Y.D.); xuhao@nudt.edu.cn (H.X.); liuchangsen21@nudt.edu.cn (C.L.); zengweixin13@nudt.edu.cn (W.Z.)

\* Correspondence: jiuyang\_tang@nudt.edu.cn

**Abstract:** The inductive relation prediction of knowledge graphs, as an important research topic, aims at predicting the missing relation between unknown entities with many real-world applications. Existing approaches toward this problem mostly use enclosing subgraphs to extract the features of target nodes to make predictions; however, there is a tendency to ignore the neighboring relations outside the enclosing subgraph, thus leading to inaccurate predictions. In addition, they also neglect the rich commonsense information that can help filter out less convincing results. In order to address the above issues, this paper proposes a commonsense-guided inductive relation prediction method with a dual attention mechanism called CNIA. Specifically, in addition to the enclosing subgraph, we added the multi-hop neighboring relations of target nodes, thereby forming a neighbor-enriched subgraph where the initial embeddings are generated. Next, we obtained the subgraph representations with a dual attention (i.e., edge-aware and relation-aware) mechanism, as well as the neighboring relational path embeddings. Then, we concatenated the two embeddings before feeding them into the supervised learning model. A commonsense re-ranking mechanism was introduced to filter the results that conformed to commonsense. Extensive experiments on WN18RR, FB15k-237, and NELL995 showed that CNIA achieves better prediction results when compared to the state-of-the-art models. The results suggested that our proposed model can be considered as an effective and state-of-the-art solution for inductive relation prediction.

**Keywords:** inductive relation prediction; commonsense; dual attention; contrastive learning



**Citation:** Duan, Y.; Tang, J.; Xu, H.; Liu, C.; Zeng, W. Commonsense-Guided Inductive Relation Prediction with Dual Attention Mechanism. *Appl. Sci.* **2024**, *14*, 2044. <https://doi.org/10.3390/app14052044>

Academic Editor: Luis Javier García Villalba

Received: 9 February 2024  
Revised: 25 February 2024  
Accepted: 26 February 2024  
Published: 29 February 2024



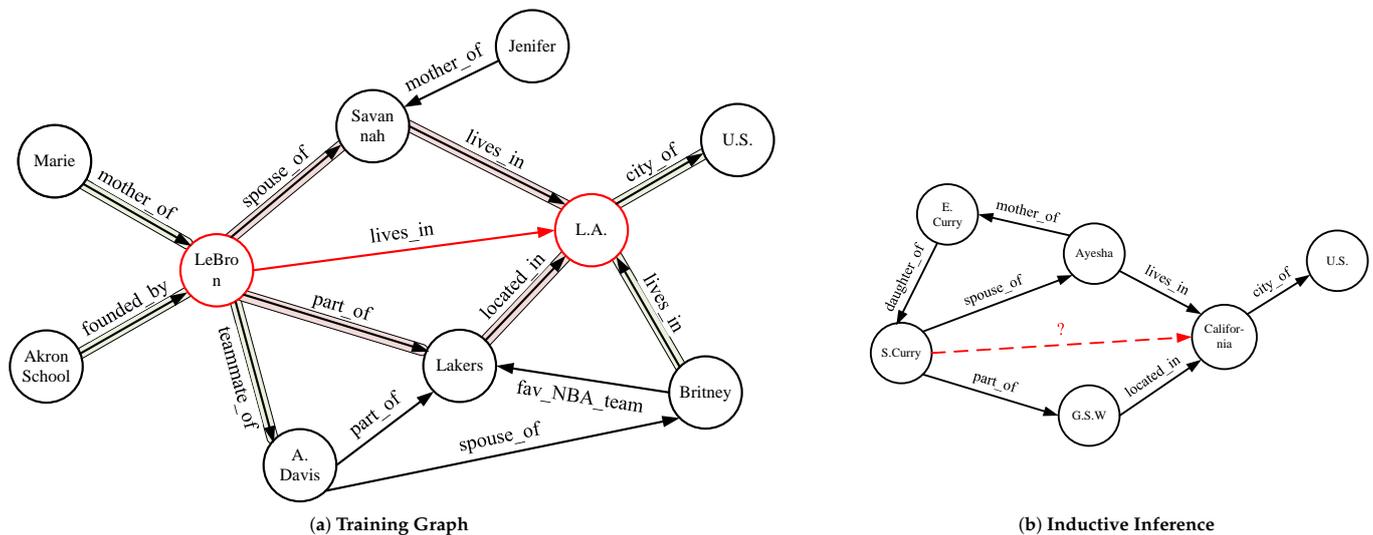
**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Knowledge graphs (KGs) are composed of organized knowledge in the form of factual triples (*entity, relation, entity*), and they form a collection of interrelated knowledge, thereby facilitating downstream tasks such as question answering [1], relation extraction [2], and recommendation systems [3]. However, even state-of-the-art KGs suffer from an issue of incompleteness [4,5], such as FreeBase [6] and WikiData [7]. To solve this issue, many studies have been proposed mining missing triples in KGs, in which the embedding-based methods become a dominant paradigm, such as TransE [8], ComplEx [9], RGCN [10], and CompGCN [11]. In particular, certain scholars have explored knowledge graph completion under low-data regime conditions [12]. In actuality, the aforementioned methods are often only suitable for *transductive* scenarios, which assumes that the set of entities in KGs is fixed.

However, KGs undergo continuous updates, whereby new entities and triples are incorporated to store additional factual knowledge, such as new users and products on e-commerce platforms. Predicting the relation links between new entities requires *inductive reasoning capabilities*, which implies that generality should be derived from existing datasets and extended to a broader spectrum of fields, as shown in Figure 1. The crux of the inductive relation prediction [13] resides in utilizing information that is not specifically tied to a particular entity. A representative strategy in inductive relation prediction techniques

is rule mining [14], which extracts first-order logic rules from a given KG and employs weighted combinations of these rules for inference. Each rule can be regarded as a relational path, comprising a set of relations from the head entity to the tail entity, which signifies the presence of a target relationship between two entities. For example, consider the straightforward rule  $(X, \text{part\_of}, Y) \wedge (Y, \text{located\_in}, Z) \rightarrow (X, \text{lives\_in}, Z)$ , which was derived from the KG depicted in Figure 1a. These relational paths exist in symbolic forms and are independent of particular entities, thus rendering them inductive and highly interpretable.



**Figure 1.** An explanatory case in inductive relation prediction, which learned from a (a) training graph, and generalizes to be (b) without any shared entities for inference. A red dashed line denotes the relation to be predicted.

Motivated by graph neural networks (GNNs) that have the ability of aggregating local information, researchers have recently proposed GNN-based inductive models. GraIL [15] models the subgraphs of target triples to capture topologies. Based on GraIL, some works [16–18] have further utilized enclosing subgraphs for inductive prediction. Recent research has also considered few-shot settings for handling unseen entities [19,20]. SNRI [21] extracts the neighbor features of target node and path features, solves the problem of sparse subgraphs, and introduces mutual information (MI) maximization to model from a global perspective, which improves the prediction effect of inductive relationships.

Nonetheless, we still observe several issues from the existing literature: (1) *Ignorance of neighboring information outside the enclosing subgraph*. Inductive relation prediction models relying on the enclosing subgraph disregard the remaining neighbors of the node. However, neighbors in proximity to the target nodes harbor valuable information for inferring the relation; (2) *Overlooking the influence of different neighboring relations*. Existing methods overlook variations in the impact of distinct structures on subgraph modeling, as well as in the diverse relationships of target nodes regarding relation predicting; and (3) *Generation of facts that violate commonsense*. Some of the predictions generated violate commonsense, which can be easily avoided by introducing the commonsense knowledge for screening.

To address these issues, we propose a Commonsense-guided Neighboring relation InfoMax model based on a dual-Attention mechanism (CNIA). CNIA was built upon the popular inductive relation prediction framework but improves on the following aspects: (1) In order to fully leverage the neighboring relations of target nodes, we constructed a *neighbor-enriched subgraph* so as to include the useful neighboring information in addition to the enclosing subgraph; (2) To account for the structure of KGs, a dual attention mechanism was employed. This mechanism combined edge-aware attention, which addresses variations in the influence of different edges on target nodes, as well as relation-aware attention, which addresses the variations in the influence of diverse relationships on predicted rela-

tionships. Together, these two types of attention enable more accurate predictions within KGs. (3) A commonsense-based re-ranking strategy was used to update the score function of the triples to filter out less convincing prediction results. By integrating the aforementioned strategies during training, CNIA ensures the complete and effective retention of neighboring relation and path information, thereby enhancing the accuracy and rationality of KG relation predictions.

**Contributions.** The contributions of this work can be summarized into three aspects:

- We put forward a commonsense-guided inductive relation prediction method with a dual attention mechanism, CNIA, which can enhance the representation learning and the accuracy of results;
- We propose to construct a neighbor-enriched subgraph to retain more useful neighboring information to aid the prediction;
- We compare the CNIA with state-of-the-art models on benchmark datasets, and the results demonstrate the superior performance of our model.

**Organization.** Section 2 provides an overview of the related works. Section 3 introduces the module. Section 4 describes the experimental design and analyzes the results, and this is followed by the conclusions in Section 5.

## 2. Related Works

In this section, we first present the existing solutions to inductive relation predictions. Next, we briefly introduce commonsense knowledge.

### 2.1. Relation Prediction Methods

In order to increase the completeness of KGs, state-of-the-art methods use a KG internal structure [8,15] or external KG information [22,23]. In this work, we focused on the former and studied the relation prediction problem.

**Transductive methods.** Transductive methods are used to learn an entity-specific embedding for each node, which have one thing in common: reasoning over original KGs. However, it is difficult to predict the missing links between unseen nodes. For example, TransE [8] is based on translation while RGCN [10] and CompGCN [11] are based on GNN. The major differences between them are the scoring function and whether the structure information in KGs is utilized. Wang et al. proposed to employ a global neighborhood aggregator to obtain the global structural information of an entity to solve the problem of sparse local structural information under certain snapshots [24]. Meng et al. proposed a multi-hop path inference model based on sparse temporal knowledge graphs [25]. Recently, Wang et al. proposed knowledge graph completion for multi-level interactions, where entities and relations interact simultaneously at both fine- and coarse-grain levels [26].

**Inductive methods.** Inductive methods can be used to learn how to reason on unseen nodes. There are two main categories of methods: rule-based and graph-based. Rule-based methods aim to learn entity-independent logical rules of reasoning. For instance, NeuralLP [14] and DRUM [27] integrate neural networks with symbolic rules to learn logic rules and rule confidence in an end-to-end microscopic manner.

Concerning graph-based methods, in recent years, researchers have drawn inspiration from the local information aggregation capability of graph neural networks, and they have incorporated graph neural networks (GNNs) into their models. GraIL [15] acquires the topology of target nodes by extracting the enclosing subgraph of the target triad, thus exhibiting inductive prediction abilities. Building on this model, TACT [16] introduces the correlation of relations in subgraphs and constructs a relational correlation network (RCN) to enhance the encoding of subgraphs. CoMPLIE [17] proposes a node-edge communication message propagation network to enhance the interaction between nodes and edges, and it naturally deals with asymmetric or anti-symmetric relations so as to enhance the adequate flow of relational information. ConGLR [13] formulates a contextual graph that represents the relationship paths in the subgraph, whereby two GCNs are applied to deal with the

enclosing subgraphs and contexts, in which different layers utilize counterpart outputs interactively for better feature representation. RE-PORT [28] aggregates relational paths and contexts to capture the connections and intrinsic nature of entities through a unified hierarchical transformation framework. However, since the experimental metrics of RE-PORT are different from those of other state-of-the-art models, the RE-PORT model was not chosen as a comparison model in the experimental part of this paper. RMPI [29] uses a novel relational message passing network for fully inductive knowledge graph completion. SNRI [21] extracts the neighbor relational features and path embeddings of nodes to make full use of the full neighbor relational information of the entities to obtain a better inductive effect. However, all these methods only add additional simple processing load, and they do not take full advantage of the whole structural feature of KGs. Different from SNRI, CNIA keeps integrated neighboring relations, utilizes a dual attention mechanism to process the structural feature of subgraphs, and it introduces commonsense re-ranking.

## 2.2. Commonsense Knowledge

Commonsense knowledge is a crucial link in addressing the bottlenecks of AI and knowledge engineering technology. And the acquisition of commonsense knowledge is the basic problem in this field. The earliest construction method involves experts manually defining the schema and relationship types of the knowledge base. Lenat [30] constructed CYC, one of the oldest knowledge bases, in the 1980s. However, the expert construction method requires substantial human and material resources. Consequently, researchers have began developing semi-structured and unstructured text extraction methods. The commonsense knowledge base constructed by YAGO [31] contains more than 1 million entities and 5 million facts, which were derived from semi-structured Wikipedia data and harmonized with WordNet through a well-designed combination of rule-based and heuristic approaches.

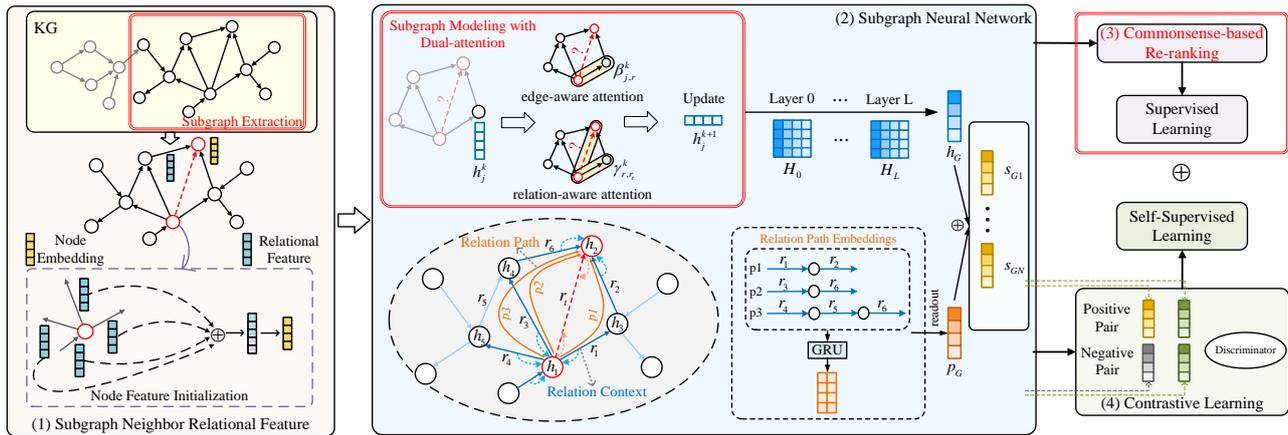
The aforementioned approaches prioritize encyclopedic knowledge and structure storage by establishing a well-defined entity space and a corresponding relationship system. However, actual general knowledge is much more loosely structured and challenging to apply to the model of two entities with known relationships. Therefore, the existing solution is to model the entity parts as natural language phrases and the relations as any concepts that can connect the entities. The OpenIE approach, for example, reveals the properties of open-text entities and relations. However, the method is extractive and it is difficult to obtain the semantic information of the text.

## 3. Methodology

In this section, we introduce the framework of CNIA. Next, we describe the foundation framework of the inductive relation prediction. Finally, we elaborate on the components of CNIA. An overview of our proposed model, CNIA, is shown in Figure 2.

### 3.1. Problem Statement

The inductive relationship prediction aims to predict the relationships between entities outside the training set. Given the KG dataset and the target triple as inputs, the model needs to generate scores of the predicted relationships of the target triple as the outputs. Previous solutions toward this task fail to take into account the full adjacency features and commonsense constraints. To fill in these gaps, our proposed model CNIA extracts more comprehensive neighboring relational features. In addition, we applied a dual attention mechanism to improve the representation learning. We also adopt commonsense re-ranking, for the first time to our knowledge, to refine the prediction results.



**Figure 2.** The overall framework of CNIA, which consists of the following parts: (1) extraction of the subgraphs with complete neighboring relations and initialization of the node features; (2) the feeding of subgraphs into the SNN with dual attention mechanisms to learn representations and extract neighboring relational path features; (3) commonsense-based re-ranking; (4) maximizing the MI between subgraph to graph to model the neighboring relations in a global way.

### 3.2. Model Overview

The core idea of CNIA is that the subgraph structure and relationship information around two nodes can predict the relationship  $r_t$  between two nodes  $(u, v)$ . Thus, only the structure and relationship information is extracted as the model input. Firstly, we extracted the neighboring relational features from the multi-hop neighbor subgraph surrounding the target nodes, which resulted in the creation of a neighbor-enhanced subgraph. We then utilized the subgraph neural network (SNN) to obtain the local subgraph embeddings from this enhanced subgraph. To capture the distinct influences of the different edges on target node embeddings and the different relationships associated with the target relations, we introduced edge-aware and relation-aware attention mechanisms. These mechanisms enabled us to learn more informative representations. Additionally, we modeled the neighboring relational paths to obtain relational path embeddings, which further contribute to the overall representation learning process. Subsequently, we concatenated the subgraph embeddings and relational path embeddings before inputting them into the supervised learning phase. And we utilized MI maximization to contrast the learning from global and local perspectives. Furthermore, we introduced a commonsense re-ranking mechanism to prioritize the results that align with commonsense knowledge and conform to general expectations. This mechanism ensures the selection of more reliable and sensible predictions. Figure 2 shows the steps of the whole process.

The notation table can be found in Table 1.

### 3.3. Foundation Framework

In this subsection, we introduce the general model of inductive relation predictions.

**Neighboring Relational Feature Extraction.** The neighboring relational feature model consists of two parts: subgraph extraction and node initialization. The local graph neighborhood of a specific triple in the KGs contained the logical evidence needed to infer the relationship between the target nodes; as such, the first step was to extract the subgraphs around the target nodes that contain complete neighbor relationships to initialize the node features. For the subgraph extraction, the target triple  $(u, r_t, v)$  was identified and the enclosing subgraph  $G(u, r_t, v)$  around the target triple was extracted. More details can be found in the GraIL paper [15]. Then, regarding node initialization—considering that the inductive inference method cannot utilize the node attribute features—the node initial features were obtained by extracting the node’s positional features and adjacency features. The details can be found in [21].

**Subgraph Representaton Learning.** As the main component of the foundation framework, this stage includes two parts: (1) obtaining the representation of subgraph entity nodes by the subgraph neural networks; (2) extracting and modeling the neighboring relational paths across target triples.

**Table 1.** Definitions of the notations.

Symbols	Descriptions
$(u, r_t, v)$	Target triple
$r_t$	Target relationship to be predicted
$G(u, r_t, v)$	Extracted subgraph for target triple
$\mathbf{h}_G$	The representation of subgraph $G$
$V_G$	Node set of subgraph $G$
$\lambda$	The weight of MI maximization mechanism
$\mathbf{p}_G$	The relational path representation of the subgraph
$\mathbf{s}_G$	The representation of subgraph $G$
$\beta_{u,r}^k$	Edge-aware attention
$\gamma_{r,r_t}^k$	Relation-aware attention
$\alpha_{u,r_t,v}^k$	Joint neighbor attention
$(c_h, r_t, c_t)$	Common sense
$\omega$	The weight of commonsense in score function

For subgraph modeling, the subgraph  $G(u, r_t, v)$  of the target triad  $(u, r_t, v)$  was first input into the GNN. SNRI defines the update strategy of the node features at each layer of the neural network. Inspired by CoMPLE [17], it feeds all node embeddings  $\mathbf{H}_L$  of the last layer to a gated recurrent unit to increase the expressive ability of the network. Finally, to obtain the representation of subgraph  $G$ , it uses an average readout function as follows:

$$\mathbf{h}_G = \frac{1}{|V_G|} \sum_{i \in V_G} \mathbf{h}_i^L. \quad (1)$$

The neighboring relational path modeling aims to obtain valuable information from the relationship paths of  $(u, v)$  as this helps to predict the type of relationship between  $u$  and  $v$ . The specific procedure can be found in reference [21].

The subgraph representation  $\mathbf{h}_G$  and neighboring relational path representation  $\mathbf{p}_G$  were concatenated as the final representation of subgraph  $\mathbf{s}_G$  as follows:

$$\mathbf{s}_G = [\mathbf{h}_G \oplus \mathbf{p}_G]. \quad (2)$$

The score function of target triple is

$$\text{Score}(u, r_t, v) = \mathbf{W}_s [\mathbf{h}_u^L \oplus \mathbf{h}_v^L \oplus \mathbf{e}_{r_t}^L \oplus \mathbf{s}_G], \quad (3)$$

where  $\mathbf{h}_u^L$ ,  $\mathbf{h}_v^L$ , and  $\mathbf{e}_{r_t}^L$  represent the embeddings of the target nodes  $u$ ,  $v$ , and target relation  $r_t$  at the  $L$  layer of the SNN.

**Supervised training.** A loss function for supervised learning was constructed:

$$L_{\text{sup}} = \sum_{(u, r_t, v) \in G} \max(0, \text{Score}(u', r'_t, v') - \text{Score}(u, r_t, v) + \theta), \quad (4)$$

where  $(u, r_t, v)$  and  $(u', r'_t, v')$  refer to the positive and negative samples, and  $\theta$  is the margin hyperparameter.

**Contrastive Learning.** Contrastive learning was widely used in unsupervised learning to attract positive samples and repel negative samples to obtain a higher-quality representation. To avoid the SNN in SNRI over-emphasizing the local structure, we further modeled the neighboring relations in a global way by subgraph–graph mutual information maximization, that is, SNRI sought to enable the neighboring relational features and paths to capture the global information of the entire KG, which would be realized in [21]. And the loss function  $L_{MI}$  could also be obtained.

**Joint Training.** The ultimate learning goal of this task was to train the following loss function:

$$L = L_{\text{sup}} + \lambda L_{MI}, \quad (5)$$

where  $\lambda$  controls the contribution of the MI max mechanism.

### 3.4. Our Model

Next, we introduce our model (cf. Figure 2), which improves over the foundation framework in four respects, i.e., neighbor-enriched subgraph extraction, neighboring relational paths, subgraph modeling based on a dual attention mechanism, and common-sense re-ranking.

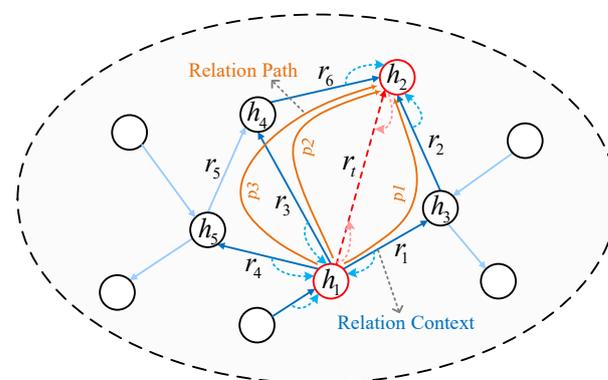
#### 3.4.1. Neighbor-Enriched Subgraph Extraction

Based on the enclosing subgraph extraction, the other neighboring relationships of each node were expanded to obtain the neighbor-enriched subgraph. The specific steps are as follows:

- Step 1 Obtain the set of three-hop neighbor nodes  $N_k(u)$  and  $N_k(v)$  of the target nodes  $u$  and  $v$  in the KG, respectively. The neighbor nodes obtained here do not distinguish the direction.
- Step 2 Take the intersection  $N_k(u) \cap N_k(v)$  of the neighbor nodes of  $u$  and  $v$  to obtain the nodes of the enclosing subgraph.
- Step 3 Filter out the isolated nodes or nodes with a distance that is greater than three from any of the target nodes to obtain the enclosing subgraph with a path length that does not exceed the length between the target nodes.
- Step 4 Keep the complete three-hop neighbor relationship  $N^r(i)$  of each node  $i$ , which includes the part omitted by the enclosing subgraph.

#### 3.4.2. Neighboring Relational Path

As shown in Figure 3, to solve the sparse subgraph problem, we modeled the neighboring relational paths.



**Figure 3.** An example of the neighbor relational paths, which are denoted by orange arrows, connecting the head to tail. And the relation context is denoted by blue edges.

A relational path is a sequence of relations on a target node, e.g.,  $p_1 = (r_1, r_2)$ , where  $r_1$  and  $r_2$  are the surrounding relations on the target nodes  $u$  and  $v$ . Define  $P$  as the set of

all neighboring relational paths on  $(u, v)$  in the subgraph. For each relation path  $p$ , GRU modeling is first applied as follows:

$$\mathbf{m}_u^k = \sum_{i \in N(u)} \mathbf{h}_i^k, \tag{6}$$

$$\mathbf{h}_i^{k+1} = \sigma\left(\left[\mathbf{m}_u^k \oplus \mathbf{m}_v^k \oplus \mathbf{h}_i^k\right] \cdot \mathbf{W}^k + \mathbf{b}^k\right), \tag{7}$$

where  $i \in N(u)$  is a neighboring edge of node  $u$  and  $\mathbf{h}_i^0$  is the initial characterization of  $i$ . Repeat the above equation for  $K$  rounds, and then splice to obtain the path representation of the subgraph as follows:

$$\mathbf{p}_G = \sigma\left(\left[\mathbf{m}_u^{K-1} \oplus \mathbf{m}_v^{K-1}\right] \cdot \mathbf{W}^{K-1} + \mathbf{b}^{K-1}\right). \tag{8}$$

### 3.4.3. Subgraph Modeling Based on a Dual Attention Mechanism

For subgraph modeling, in order to improve the accuracy of the subgraph information utilization and distinguish the role played by different structures in modeling, we propose to integrate edge and relation dual attention mechanisms into a GCN. Generally, the entity embedding for updating subgraphs employs the traditional iterative message passing strategy of a GCN.

$$\mathbf{h}_v^{k+1} = \sum_{(u,r) \in N_s(j)} \alpha_{(u,r,v)}^k \mathbf{W}_{t1}^k \phi(\mathbf{h}_u^k, \mathbf{e}_r^k) + \mathbf{W}_{t2}^k \mathbf{h}_v^k, \tag{9}$$

where  $(u, r, v)$  is the example triple and  $N_s(\cdot)$  is the set of neighbors containing entity relationship pairs.  $\mathbf{W}_{t1}^k$  and  $\mathbf{W}_{t2}^k$  are transformation matrices, and  $\mathbf{h}_v^k$  represents the embedding of entity  $v$  at layer  $k$ .

In actuality, there are differences between the neighbors in entity modeling. For example, neighbors  $(h_3, r_2)$ , as well as  $(h_4, r_6)$ , have different effects on the representation of  $h_2$ . Meanwhile, predicting different target relationships affects the representation of  $h_2$ . In order to distinguish these effects, we propose two types of attention, edge-aware  $\beta_{u,r}^k$  and relation-aware  $\gamma_{r,r_t}^k$ , to compute the joint neighbor attention values.

$$\beta_{u,r}^k = \sigma\left(\mathbf{W}_1^k \left(\mathbf{h}_u^k \oplus \mathbf{e}_r^k \oplus \mathbf{h}_v^k\right)\right), \tag{10}$$

$$\gamma_{r,r_t}^k = \sigma\left(\mathbf{W}_2^k \left(\mathbf{e}_r^k \oplus \mathbf{e}_{r_t}^k\right)\right), \tag{11}$$

$$\alpha_{(u,r,v)}^k = \text{softmax}\left(\beta_{u,r}^k + \gamma_{r,r_t}^k\right), \tag{12}$$

where  $\mathbf{W}_1^k$  and  $\mathbf{W}_2^k$  are transformation matrices.

To enable information sharing and improve representation, we integrated updated entity embeddings to enhance the representation as follows:

$$\mathbf{H}_{index(v)}^{k+1} = \lambda_1 \mathbf{H}_{index(v)}^k + (1 - \lambda_1) \mathbf{h}_v^{k+1}, \tag{13}$$

where  $index(v)$  indicates the actual index of entity  $v$ .

Based on this, we can obtain the embedding matrix  $\mathbf{H}_L$ , which consists of the embedding vectors of all the entities in the subgraph of the last GCN. To further improve the representational power of the model, a two-way gated recursive unit GRU, as inspired by [17,32], was added after the last layer of the GCN. Since the entities in the subgraph are sorted according to the distances to the target entities, the GRU that handles the sequences can be used to increase the inter-entity interactions, thereby updating the embedding representation as follows:

$$\mathbf{H}_L' = \text{BiGRU}(\mathbf{H}_L). \tag{14}$$

Finally, the average readout function (1) was used to process  $\mathbf{H}_L'$  to obtain the representation  $\mathbf{h}_G$  of the subgraph  $G$ .

#### 3.4.4. Commonsense Re-Ranking

To ensure that the generated triples conform to commonsense, they are compared to commonsense knowledge. The generation of commonsense knowledge is specifically divided into four steps:

**Fine tuning BERT to acquire contextual representations.** The pre-trained language model BERT is suitable for acquiring the contextual representations of triples. Specifically, given a tuple  $(u, r, v)$ ,  $u$ ,  $r$ , and  $v$  are iteratively masked such that the encoder predicts the masked elements using the other two unmasked elements, which enables the BERT model to better understand the relationships between the tuple elements.

**Filtering abstract concepts.** Replacing ternary entities with entity concepts may result in concepts with a high level of abstraction. We use entity and concept representations to compute the probability of a concept appearing in an entity to measure the abstraction level of the concept. A threshold is set to filter out concepts with a higher abstraction level whose occurrence expectation is lower than the threshold.

**Entity-to-concept mapping.** After filtering out the higher degree of abstraction, concepts were used instead of entities in the triad to obtain conceptual triples. Commonsense knowledge in the individual form C1 was obtained by eliminating duplicate concept-level triples. Commonsense knowledge in the set form C2 was then obtained by merging concept-level triples that contained the same relations.

**Filtering relationship-independent concepts.** The commonsense knowledge obtained by substituting concepts for entities would also suffer from the problem that concept  $c$  is not related to relation  $e_r$ . To measure the degree of relevance of concept  $c$  and relation  $e_r$ , the following cosine similarity was calculated to obtain the similarity score.

$$R(r, c) = \cos(\mathbf{e}_r, \mathbf{c}), \quad (15)$$

$$\mathbf{e}_r = \text{mean}_{\mathbf{e}_i \in N_r}(\mathbf{e}_i), \quad (16)$$

where  $N_r$  denotes the set of relations  $r$  corresponding to the head entity or the tail entity, and  $\mathbf{c}$  and  $\mathbf{e}_r$  are the contextual representations of concepts and entities, respectively. A threshold was set to filter the target of relationship-independent concepts.

Based on the constructed commonsense knowledge database, we propose a simple commonsense-based re-ranking strategy, i.e., one that is based on the KGs of the train set or test set. Moreover, if the triple  $(u, r_t, v)$  satisfies the commonsense  $(c_h, r_t, c_t)$ , the score of the ternary is increased by  $\omega > 0$ .

$$\text{Score}(u, r_t, v) = \mathbf{W}_s \left[ \mathbf{h}_u^L \oplus \mathbf{h}_v^L \oplus \mathbf{e}_{r_t}^L \oplus \mathbf{s}_G \right] + \omega (I((u, r_t, v) \in (c_h, r_t, c_t))), \quad (17)$$

where  $I(\cdot)$  is the indicator function. We substituted the updated triple score function into the original loss function (4). Then, we used MI maximization to obtain a global representation and train joint training strategy (5).

#### 3.4.5. Algorithmic Descriptions

The algorithm of the whole model is presented in Algorithm 1.

## 4. Experiments

In this section, we first introduce the experimental configurations. Next, we provide the empirical results and discuss the performance, and then we finally discuss the further experiments that were conducted.

#### 4.1. Experimental Configurations

In this subsection, we introduce the experimental configurations, including the datasets, evaluation metrics, parameter settings, and the baseline models.

##### 4.1.1. Datasets

WN18RR [33], FB18k-237 [34], and NELL-995 [35] are widely used in transductive link prediction. As for inductive relation prediction, we follow GraIL to conduct experiments based on the variants of WN18RR, FB18k-237, and NELL-995, where entities in a test set are not contained in a train set. And every dataset induces four versions with increasing sizes.

---

#### Algorithm 1 The inductive process of the CNIA model

---

**Input:**  $KG$ , target triple  $(u, r_t, v)$

**Output:** the *Score* of  $(u, r_t, v)$

- 1: Extract subgraph  $G$  of target triple  $(u, r_t, v)$  from  $KG$
  - 2: Initialize node representation  $\mathbf{h}$  for all the nodes in  $G$
  - 3: **for all**  $k$  in  $K$  of GCN layers **do**
  - 4:   Update the embedding  $\mathbf{h}_j^k$  of each entity  $j$  at layer  $k$
  - 5:   Calculate edge attention value  $\beta_{r,r_t}^k$  according to Equation (10)
  - 6:   Calculate relation attention value  $\gamma_{r,r_t}^k$  according to Equation (11)
  - 7:   Generate general attention  $\alpha_{(u,r,v)}^k$  according to Equation (12)
  - 8:   Update the embedding matrix  $\mathbf{H}_{\text{index}(v)}^k$  of all entities
  - 9: **end for**
  - 10: Obtain entity embedding  $\mathbf{H}_L'$  through BiGRU
  - 11: Obtain  $\mathbf{h}_G$  through a readout function
  - 12: Establish a Neighboring Relational Path Model to obtain path representation  $\mathbf{p}_G$  Equation (8)
  - 13: Establish score function *Score* and loss function  $L_{sup}$  by Equations (3) and (4)
  - 14: Update *Score* by commonsense re-ranking according to Equation (17)
  - 15: **return** *Score*
- 

The specific sampling method was divided into the following three steps: (1) Uniformly sample several root entities as root nodes in the original KGs; (2) Take the concatenated set of the multi-hop neighbors of the root node as the training KGs. To prevent the number of nodes from growing exponentially, the size of  $k$  needs to be limited; (3) Remove the sampled training set from the entire original KGs and then repeat Steps 1-2. Then, sample the test set so as to ensure that the entities in the test set are not included in the training set. The detailed statistics of the dataset are shown in Table 2.

**Table 2.** The statistics of three inductive datasets. #E and #R, respectively, denote the number of entities and relations. #TR1 represents triples, which are utilized to form KG  $G$ , while triples in #TR2 are utilized to evaluate.

Version	Split	#R	WN18RR			FB15k-237			NELL-995				
			#E	#TR1	#TR2	#R	#E	#TR1	#TR2	#R	#E	#TR1	#TR2
v1	Train	9	2746	5410	630	183	2000	4245	489	14	10,915	4687	414
	Test	9	922	1618	188	146	1500	1993	205	14	225	833	100
v2	Train	10	6954	15,262	1838	203	3000	9739	1166	88	2564	8219	922
	Test	10	2923	4011	441	176	2000	4145	478	79	4937	4586	476
v3	Train	11	12,078	25,901	3097	218	4000	17,986	2194	142	4647	16,393	1851
	Test	11	5084	6327	605	187	3000	7406	865	122	4921	8048	809
v4	Train	9	3861	7940	934	222	5000	27,203	3352	77	2092	7546	876
	Test	9	7208	12,334	1429	204	3500	11,714	1424	61	3294	7073	731

#### 4.1.2. Evaluation Metrics

The majority of existing models, such as SNRI, use AUC-PR and Hits@10 as evaluation metrics. In order to make a fair comparison, we followed previous works and also used these two metrics. Specifically, AUC-PR is a classification indicator that computes the area under the precision–recall curve. Hits@10 is defined as the proportion of correct entities ranked in a top 10 of candidate entities. The results are obtained by averaging over five runs for an accurate evaluation.

#### 4.1.3. Parameter Settings

In the subgraph extraction section, the experiment selected  $k = 3$ , indicating that only the neighboring nodes within three hops of the target node were retained for the union operation. The learning rate was set to 0.0005, the dropout rate was 0.5, and the embedding dimension was 32. The margin parameter  $\theta$  for the supervised learning loss function was 10, and the coefficient  $\lambda$  for the joint loss function was 5. Simultaneously, L2 regularization was applied to prevent overfitting, the maximum training epoch was set to 50, and Adam was employed as the optimizer for the training model parameters. All experiments were conducted in PyTorch and executed on an NVIDIA RTX 3090.

#### 4.1.4. Baseline Models

To explore the performance of our proposed model, we compared five state-of-the-art GCN-based baselines, including the following:

- **GraIL** [15]: This method pioneered a novel approach to inductive reasoning by introducing subgraph encoding for the first time, whereby it addressed the invisible entities in entirely new KGs.
- **TACT** [16]: This approach models the semantics between relationships and uses relationship topology to detect correlations for inductive relation predictions.
- **CoMPiLE** [17]: This method is grounded in the structure of local directed subgraphs, and it exhibits a robust inductive bias for handling entity-independent semantic relations.
- **SNRI** [21]: This approach leverages the full adjacency of entities in subgraphs using neighbor relationship features and neighbor relationship paths. This forms the basis for inductive relation predictions in KGs.
- **ConGLR** [13]: This method constructs a contextual graph that represents relational paths and subsequently processes them with two GCNs, each incorporating enclosing subgraphs.
- **RMPI** [29]: This approach passes messages directly between relations to make full use of the relation patterns for subgraph reasoning.

### 4.2. Experimental Results

In this subsection, we first report the main results. Then, we discuss the ablation study that was subsequent conducted.

#### 4.2.1. Main Results

The main experimental results are presented in Tables 3 and 4. “Avg.” denotes the average metric values of the four versions of the same KG. The optimal and suboptimal results are marked in bold and underlined text, respectively. “—” represents the results that were not provided in the original work (and cannot be reproduced).

**Table 3.** The AUC-PR metric values (%) of inductive relation predictions on twelve dataset versions.

Model	WN18RR					FB15k-237					NELL-995				
	v1	v2	v3	v4	Avg.	v1	v2	v3	v4	Avg.	v1	v2	v3	v4	Avg.
GraIL	94.32	94.18	85.80	92.72	91.75	84.69	90.57	91.68	94.46	90.35	86.05	92.62	93.34	87.50	89.87
TACT	95.43	97.54	87.65	96.04	94.16	83.15	93.01	92.10	94.25	90.62	81.06	93.12	96.07	85.75	89.00
CoMPILE	98.23	99.56	93.60	99.80	97.79	85.50	91.68	93.12	<u>94.90</u>	91.30	80.16	<b>95.88</b>	96.08	85.48	89.04
SNRI	99.10	<b>99.92</b>	94.90	99.61	<u>98.38</u>	<u>86.69</u>	91.77	91.22	93.37	90.76	—	—	—	—	—
ConGLR	<u>99.58</u>	99.67	93.78	<u>99.88</u>	98.22	85.68	<u>92.32</u>	<b>93.91</b>	<b>95.05</b>	<u>91.74</u>	<u>86.48</u>	95.22	<u>96.16</u>	<u>88.46</u>	<u>91.58</u>
RMPI	95.00	95.96	88.53	95.78	93.82	85.25	92.19	92.09	92.80	90.58	81.12	93.46	95.35	<b>91.77</b>	90.43
CNIA (ours)	<b>99.89</b>	<u>99.91</u>	<b>94.95</b>	<b>99.90</b>	<b>98.66</b>	<b>86.99</b>	<b>92.75</b>	<u>93.42</u>	94.80	<b>91.99</b>	<b>95.50</b>	<u>95.43</u>	<b>96.25</b>	88.13	<b>93.83</b>

**Table 4.** The Hits@10 metric values (%) of inductive relation predictions on twelve dataset versions.

Model	WN18RR					FB15k-237					NELL-995				
	v1	v2	v3	v4	Avg.	v1	v2	v3	v4	Avg.	v1	v2	v3	v4	Avg.
GraIL	82.45	78.68	58.43	73.41	73.24	64.15	81.80	82.83	89.29	79.51	59.50	93.25	91.41	73.19	79.33
TACT	84.04	81.63	67.97	76.56	77.55	65.76	83.56	85.20	88.69	80.80	79.80	88.91	94.02	73.78	84.12
CoMPILE	83.60	79.82	60.69	75.49	74.90	67.64	82.98	84.67	87.44	80.68	58.38	93.87	92.77	75.19	80.05
SNRI	<u>87.23</u>	83.10	67.31	83.32	80.24	<u>71.79</u>	<u>86.50</u>	<b>89.59</b>	<u>89.39</u>	84.32	—	—	—	—	—
ConGLR	85.64	<b>92.93</b>	<u>70.74</u>	<b>92.90</b>	<b>85.55</b>	68.29	85.98	88.61	89.31	<u>82.93</u>	<u>81.07</u>	<b>94.92</b>	<u>94.36</u>	81.61	<u>87.99</u>
RMPI	82.45	78.68	58.68	73.41	73.31	65.37	81.80	81.10	87.25	78.88	59.50	92.23	93.57	<b>87.62</b>	83.23
CNIA (ours)	<b>89.36</b>	<u>85.94</u>	<b>72.06</b>	<u>83.03</u>	<u>82.60</u>	<b>74.20</b>	<b>86.51</b>	<u>89.43</u>	<b>89.39</b>	<b>84.88</b>	<b>85.22</b>	<u>94.11</u>	<b>94.55</b>	<u>82.03</u>	<b>88.98</b>

The two tables present the results of the comparative experiments. CNIA demonstrated optimal performance in both AUC-PR and Hits@10 metrics across most datasets, thus confirming the effectiveness and sophistication of CNIA. Specifically, the CNIA model outperformed the GraIL, TACT, and CoMPILE models on all datasets, and it surpassed the SNRI and ConGLR models on most datasets, albeit with suboptimal results on individual datasets. When averaging the metrics across the three datasets, it was evident that the average AUC-PR value of CNIA was optimal on all three datasets, while the Hits@10 value was suboptimal on WN18RR and optimal on the other two datasets.

Compared with the base model SNRI, the experimental results on both WN18RR and FB15k-37 were significantly improved, and the average AUC-PR and Hits@10 values were greatly increased, thereby indicating that CNIA is superior and can enhance the performance of inductive link prediction. Compared with the state-of-the-art model RMPI, CNIA is constantly better on the majority of datasets.

In comparison to the ConGLR model, the average AUC-PR improved by 0.44% on the WN18RR dataset, 0.25% on the FB15k-237 dataset, and 2.25% on the NELL-95 dataset. However, the average Hits@10 was lower than ConGLR on the WN18RR dataset, but it improved by 1.95% on the FB15k-237 dataset and 0.99% on the NELL-95 dataset. ConGLR only performed well on WN18RR-v2,v4 as on other datasets it did not perform well, which indicates that ConGLR's biased logical reasoning method is specifically suited for certain WN18RR datasets. In contrast, our model CNIA consistently performed well across various datasets, thus demonstrating its versatility and stability across different scenarios.

Among the AUC-PR metric values, CNIA exhibited more pronounced advantages on the WN18RR and NELL-995 datasets. Upon observing Table 2, it was noted that the #R value of WN18RR and NELL-995 datasets consistently remained lower than that of FB15k-237, thus indicating that the subgraphs in these two datasets were more likely to be sparse and lack the structural information for reasoning. CNIA effectively leveraged the adjacent relationship information when compared to the baseline model, thereby addressing the issue of subgraph sparsity. However, the improvement effect on the FB15-237 dataset was not as noticeable, possibly because the subgraph density of this dataset was high and due to the fact that the effective structural information could be extracted using closed subgraphs, thus making the baseline model easily predictable.

#### 4.2.2. Ablation Study

The ablation study explored the impact of each component on the overall model performance, and we conducted the ablation experiment using the WN18RR dataset. This experiment mainly consisted of five parts: (1) removal of the dual attention mechanism (w/o Att), (2) removal of the neighboring relational feature module (w/o NRF), (3) removal of the neighboring relational path module (w/o NRP), (4) removal of the commonsense-based re-ranking module (w/o CSR), (5) removal of the comparative learning (w/o CL). Table 5 shows the results of the ablation experiment, and we were able to find that the CNIA removal of any of the modules was worse than the original model, thus proving the effectiveness of each module.

**Table 5.** The Hits@10 metric values (%) of the ablation experiment.

Ablation	WN18RR			
	v1	v2	v3	v4
CNIA	<b>89.36</b>	<b>85.94</b>	<b>72.06</b>	<b>83.03</b>
CNIA w/o Att	85.64	84.69	66.69	82.40
CNIA w/o NRF	86.45	85.06	71.40	72.98
CNIA w/o NRP	84.16	85.10	71.81	74.07
CNIA w/o CSR	86.44	83.90	70.00	82.15
CNIA w/o CL	85.12	80.95	64.49	73.79

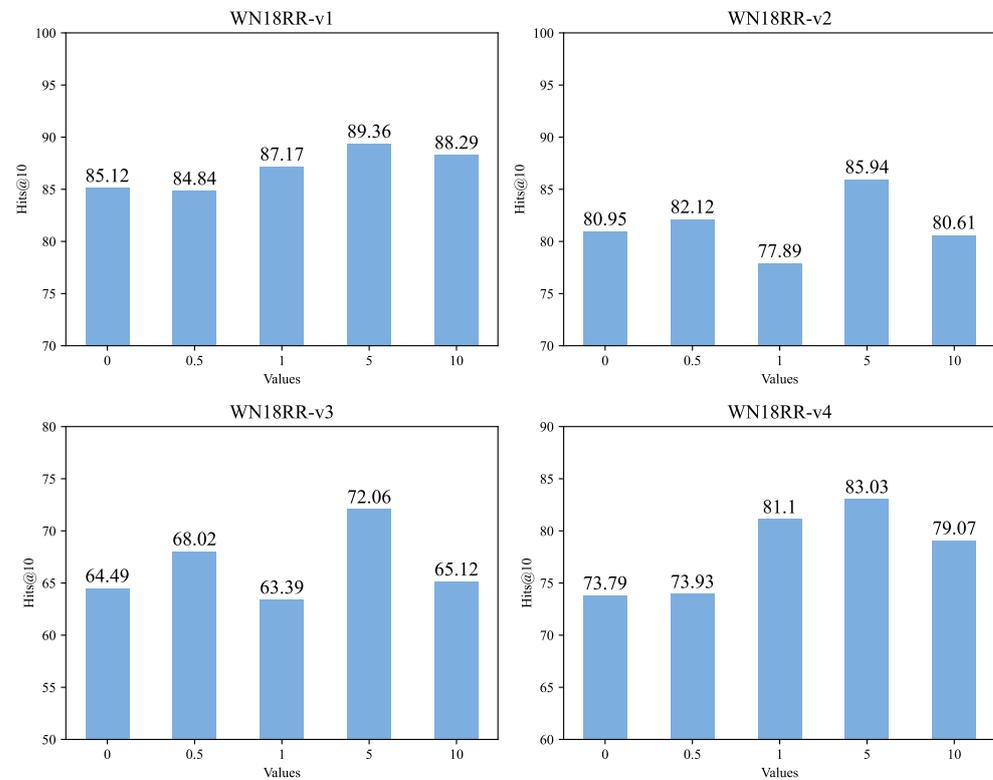
- **CNIA w/o Att:** Remove the dual attention mechanism and directly splice to obtain the neighboring relationship features. The Hits@10 was reduced by 3.72%, 1.25%, 5.37%, and 0.63%. This showed that ignoring the effect of different edges on nodes and the effect of different relations on the target relation reduces the accuracy of inference.
- **CNIA w/o NRF:** Remove the neighbor relational features and predict directly from the node features obtained by initialization. Ignoring the neighbor relationship makes the node features less expressive, which results in losing the effective information, and it also cannot completely portray the nodes.
- **CNIA w/o NRP:** Remove the neighboring relational path feature and ignore the message propagation path feature from the head node to the tail node. This makes the performance degradation obvious, thereby indicating that the neighboring relational path feature plays an important role in dealing with sparse subgraphs.
- **CNIA w/o CSR:** Remove the commonsense re-ranking module that fails to filter out the predictions that do not conform to commonsense. The experiments verified that the absence of the commonsense re-ranking will produce relationships that do not conform to commonsense, thus degrading model performance.
- **CNIA w/o CL:** Remove the contrastive learning and do not perform this operation of MI maximization. This resulted in the Hits@10 being reduced by 4.24%, 4.99%, 7.57%, and 9.24%. The results showed that, without contrastive learning, the results of most of the metrics dropped, thus demonstrating that global information helps to better model neighboring relational features.

#### 4.2.3. Hyper-Parameter $\lambda$ Analysis

In CNIA,  $\lambda$  governs the contribution of the contrastive learning MI InfoMax mechanism (Equation (5)). In this section, we further analyze the multiple values of lambda in [0, 0.5, 1.0, 5.0, 10.0]. The changes in Hits@10 values in WN18RR-v1, v2, v3, and v4 are depicted in Figure 4. It was observed that inductive relation prediction performance varies with  $\lambda$  values. Both excessively large and small  $\lambda$  values lead to a decrease in model performance.

When  $\lambda = 1$ , the results are the worst on the v2 and v3 datasets, and they are moderate on the v1 and v4 datasets, thus indicating that the prediction performance is unstable. Thus, simply setting the weights of the two to be equal is not conducive to maximizing the performance. Notably, the effect is generally better when  $\lambda > 1$  compared to  $\lambda < 1$ , thereby indicating that contrastive learning plays a more significant role than supervised learning.

In summary, this discussion establishes that a better performance can be achieved when the coefficient  $\lambda$  is set to five, and it underscored the importance of contrastive learning through the ablation experiments.



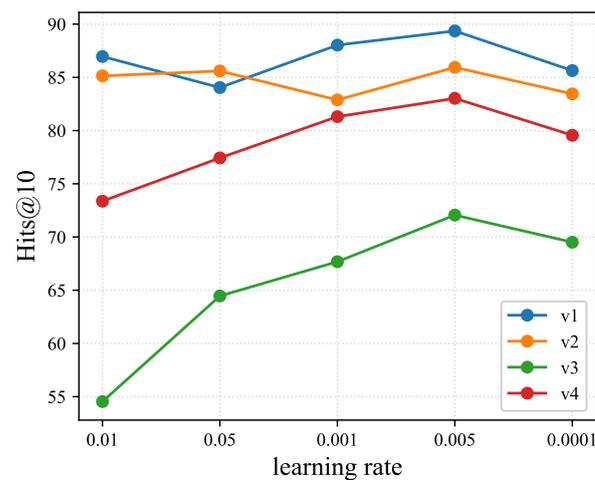
**Figure 4.** Discussion of the  $\lambda$  values in the WN18RR dataset.

#### 4.2.4. Learning Rate Analysis

The learning rate is an important hyperparameter for controlling parameter updates during neural network training. The value of the learning rate affects the training efficiency. When the learning rate is too small, the learning speed is slow, which means that it is vulnerable to overfitting and the convergence speed is slow; when the learning rate is too large, the learning speed is fast but prone to the oscillation problem such that the training value fluctuates up and down in the optimal value.

The learning rate of the basic model SNRI in the experiment was set to 0.001. This was chosen in order to improve the efficiency and effect of the experiment. Next, we discussed the learning rate of the experiment and, respectively, took the values of 0.01, 0.05, 0.001, 0.005, and 0.0001 in the WN18RR dataset to conduct the experiment.

As shown in Figure 5, the experimental results on the v1 and v2 datasets were more stable when the learning rate was changed, with the Hits@10 remaining above 80%. When the learning rate was 0.01, the experimental effect was the worst on the v3 and v4 datasets. And when the learning rate was set to 0.0001, the Hits@10 decreased on all sub-datasets, thus indicating that the learning rate was too small to affect the prediction effect. The overall trend peaked at a learning rate of 0.005 and then fell back.



**Figure 5.** Learning rate discussion on WN18RR.

## 5. Conclusions

In this study, we proposed a novel inductive relation prediction method called CNIA, which leverages commonsense knowledge and employs a dual attention mechanism to enhance relationship inference accuracy. By fully utilizing neighbor relational information and incorporating a commonsense filtering process, CNIA was able to extract neighbor relational features through an edge-aware and relation-aware dual attention mechanism. Moreover, we employed contrastive learning for global modeling to further improve the performance of our model. The experimental results on the three benchmark datasets showed that CNIA outperforms other state-of-the-art models (we conducted an ablation study to verify the effectiveness of each module).

Regarding future research prospects, reasoning with multimodal knowledge graphs could be explored, where elements such as images can be added to improve inductive reasoning. KG completion with temporal constraints is also worthy of exploration, where temporal information should be taken into account when scoring triples. Building on existing research, we plan to extend our exploration into the temporal dimension and investigate inductive relation prediction methods on time series knowledge graphs. In addition, as large language models continue to advance, there is great potential to leverage these models to enhance inductive reasoning and build more accurate inductive models.

**Author Contributions:** Conceptualization, Y.D. and C.L.; methodology, W.Z. and Y.D.; software, H.X.; validation, J.T., H.X. and W.Z.; formal analysis, Y.D.; investigation, C.L.; resources, W.Z.; data curation, H.X.; writing—original draft preparation, Y.D.; writing—review and editing, W.Z.; visualization, C.L.; supervision, H.X.; project administration, J.T.; funding acquisition, J.T. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was partially supported by NSFC under grants No. 62302513.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset can be obtained from the Github repository: <https://github.com/TimDettmers/ConvE> (accessed on 15 January 2024) and <https://github.com/xwhan/DeepPath> (accessed on 15 January 2024).

**Conflicts of Interest:** The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Zhang, Y.; Dai, H.; Kozareva, Z.; Smola, A.J.; Song, L. Variational Reasoning for Question Answering With Knowledge Graph. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI 2018), New Orleans, LA, USA, 2–7 February 2018; pp. 6069–6076.
2. Verlinden, S.; Zaporozets, K.; Deleu, J.; Demeester, T.; Develder, C. Injecting Knowledge Base Information into End-to-End Joint Entity and Relation Extraction and Coreference Resolution. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*; Association for Computational Linguistics: Kerrville, TX, USA, 2021; pp. 1952–1957.
3. Wang, H.; Zhao, M.; Xie, X.; Li, W.; Guo, M. Knowledge Graph Convolutional Networks for Recommender Systems. In Proceedings of the WWW 2019, San Francisco, CA, USA, 13–17 May 2019; pp. 3307–3313.
4. Zhao, X.; Zeng, W.; Tang, J. *Entity Alignment—Concepts, Recent Advances and Novel Approaches*; Springer: Singapore, 2023. [[CrossRef](#)]
5. Zeng, W.; Zhao, X.; Li, X.; Tang, J.; Wang, W. On entity alignment at scale. *VLDB J.* **2022**, *31*, 1009–1033. [[CrossRef](#)]
6. Bollacker, K.D.; Evans, C.; Paritosh, P.K.; Sturge, T.; Taylor, J. Freebase: A collaboratively created graph database for structuring human knowledge. In Proceedings of the SIGMOD Conference 2008, Vancouver, BC, Canada, 10–12 June 2008; pp. 1247–1250.
7. Vrandečić, D. Wikidata: A new platform for collaborative data collection. In Proceedings of the WWW 2012, Lyon, France, 16–20 April 2012; pp. 1063–1064.
8. Bordes, A.; Usunier, N.; García-Durán, A.; Weston, J.; Yakhnenko, O. Translating Embeddings for Modeling Multi-relational Data. In Proceedings of the NIPS 2013, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 2787–2795.
9. Trouillon, T.; Welbl, J.; Riedel, S.; Gaussier, É.; Bouchard, G. Complex Embeddings for Simple Link Prediction. In Proceedings of the 33rd International Conference on International Conference on Machine Learning (ICML 2016), New York, NY, USA, 19–24 June 2016; Volume 48, pp. 2071–2080.
10. Schlichtkrull, M.S.; Kipf, T.N.; Bloem, P.; van den Berg, R.; Titov, I.; Welling, M. Modeling Relational Data with Graph Convolutional Networks. In *The Semantic Web, Proceedings of the 15th International Conference, ESWC 2018, Heraklion, Crete, Greece, 3–7 June 2018; Proceedings 15*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2018; Volume 10843, pp. 593–607.
11. Vashishth, S.; Sanyal, S.; Nitin, V.; Talukdar, P.P. Composition-based Multi-Relational Graph Convolutional Networks. In Proceedings of the ICLR 2020, Addis Ababa, Ethiopia, 26–30 April 2020.
12. Liu, J.; Fan, C.; Zhou, F.; Xu, H. Complete feature learning and consistent relation modeling for few-shot knowledge graph completion. *Expert Syst. Appl.* **2024**, *238*, 121725. [[CrossRef](#)]
13. Lin, Q.; Liu, J.; Xu, F.; Pan, Y.; Zhu, Y.; Zhang, L.; Zhao, T. Incorporating Context Graph with Logical Reasoning for Inductive Relation Prediction. In Proceedings of the SIGIR 2022, Madrid, Spain, 11–15 July 2022; pp. 893–903.
14. Yang, F.; Yang, Z.; Cohen, W.W. Differentiable Learning of Logical Rules for Knowledge Base Reasoning. In Proceedings of the NIPS 2017, Long Beach, CA, USA, 4–9 December 2017; pp. 2319–2328.
15. Teru, K.K.; Denis, E.G.; Hamilton, W.L. Inductive Relation Prediction by Subgraph Reasoning. In Proceedings of the ICML 2020, Virtual, 13–18 July 2020; Volume 119, pp. 9448–9457.
16. Chen, J.; He, H.; Wu, F.; Wang, J. Topology-Aware Correlations Between Relations for Inductive Link Prediction in Knowledge Graphs. In Proceedings of the AAAI 2021, Virtual, 2–9 February 2021; pp. 6271–6278.
17. Mai, S.; Zheng, S.; Yang, Y.; Hu, H. Communicative Message Passing for Inductive Relation Reasoning. In Proceedings of the AAAI 2021, Virtual, 2–9 February 2021; pp. 4294–4302.
18. Galkin, M.; Denis, E.G.; Wu, J.; Hamilton, W.L. NodePiece: Compositional and Parameter-Efficient Representations of Large Knowledge Graphs. In Proceedings of the ICLR 2022, Virtual, 25–29 April 2022.
19. Baek, J.; Lee, D.B.; Hwang, S.J. Learning to Extrapolate Knowledge: Transductive Few-shot Out-of-Graph Link Prediction. In Proceedings of the NeurIPS 2020, Virtual, 6–12 December 2020.
20. Zhang, Y.; Wang, W.; Chen, W.; Xu, J.; Liu, A.; Zhao, L. Meta-Learning Based Hyper-Relation Feature Modeling for Out-of-Knowledge-Base Embedding. In Proceedings of the CIKM 2021, Gold Coast, QLD, Australia, 1–5 November 2021; pp. 2637–2646.
21. Xu, X.; Zhang, P.; He, Y.; Chao, C.; Yan, C. Subgraph Neighboring Relations Infomax for Inductive Link Prediction on Knowledge Graphs. In Proceedings of the IJCAI 2022, Vienna, Austria, 23–29 July 2022; pp. 2341–2347.
22. Zeng, W.; Zhao, X.; Tang, J.; Lin, X. Collective Entity Alignment via Adaptive Features. In Proceedings of the 36th IEEE International Conference on Data Engineering, ICDE 2020, Dallas, TX, USA, 20–24 April 2020; pp. 1870–1873.
23. Zeng, W.; Zhao, X.; Tang, J.; Lin, X.; Groth, P. Reinforcement Learning-based Collective Entity Alignment with Adaptive Features. *ACM Trans. Inf. Syst.* **2021**, *39*, 26:1–26:31. [[CrossRef](#)]
24. Wang, J.; Lin, X.; Huang, H.; Ke, X.; Wu, R.; You, C.; Guo, K. GLANet: Temporal knowledge graph completion based on global and local information-aware network. *Appl. Intell.* **2023**, *53*, 19285–19301. [[CrossRef](#)]
25. Meng, X.; Bai, L.; Hu, J.; Zhu, L. Multi-hop path reasoning over sparse temporal knowledge graphs based on path completion and reward shaping. *Inf. Process. Manag.* **2024**, *61*, 103605. [[CrossRef](#)]
26. Wang, J.; Wang, B.; Gao, J.; Hu, S.; Hu, Y.; Yin, B. Multi-Level Interaction Based Knowledge Graph Completion. *IEEE ACM Trans. Audio Speech Lang. Process.* **2024**, *32*, 386–396. [[CrossRef](#)]
27. Sadeghian, A.; Armandpour, M.; Ding, P.; Wang, D.Z. DRUM: End-To-End Differentiable Rule Mining on Knowledge Graphs. In Proceedings of the NeurIPS 2019, Vancouver, BC, Canada, 8–14 December 2019; pp. 15321–15331.

28. Li, J.; Wang, Q.; Mao, Z. Inductive Relation Prediction from Relational Paths and Context with Hierarchical Transformers. In Proceedings of the ICASSP 2023, Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
29. Geng, Y.; Chen, J.; Pan, J.Z.; Chen, M.; Jiang, S.; Zhang, W.; Chen, H. Relational Message Passing for Fully Inductive Knowledge Graph Completion. In Proceedings of the ICDE 2023, Anaheim, CA, USA, 3–7 April 2023; pp. 1221–1233.
30. Lenat, D.B. CYC: A Large-Scale Investment in Knowledge Infrastructure. *Commun. ACM* **1995**, *38*, 32–38. [[CrossRef](#)]
31. Suchanek, F.M.; Kasneci, G.; Weikum, G. Yago: A core of semantic knowledge. In Proceedings of the WWW 2007, Banff, AB, Canada, 8–12 May 2007; pp. 697–706.
32. Song, Y.; Zheng, S.; Niu, Z.; Fu, Z.; Lu, Y.; Yang, Y. Communicative Representation Learning on Attributed Molecular Graphs. In Proceedings of the IJCAI 2020, Yokohama, Japan, 7–15 January 2021; pp. 2831–2838.
33. Dettmers, T.; Minervini, P.; Stenetorp, P.; Riedel, S. Convolutional 2D Knowledge Graph Embeddings. In Proceedings of the AAAI 2018, New Orleans, LO, USA, 2–7 February 2018; pp. 1811–1818.
34. Toutanova, K.; Chen, D.; Pantel, P.; Poon, H.; Choudhury, P.; Gamon, M. Representing Text for Joint Embedding of Text and Knowledge Bases. In Proceedings of the EMNLP 2015, Lisbon, Portugal, 17–21 September 2015; pp. 1499–1509.
35. Xiong, W.; Hoang, T.; Wang, W.Y. DeepPath: A Reinforcement Learning Method for Knowledge Graph Reasoning. In Proceedings of the EMNLP 2017, Copenhagen, Denmark, 7–11 September 2017; pp. 564–573.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.