This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2022.3218850

OERL: Enhanced Representation Learning via Open Knowledge Graphs

Qian Li, Daling Wang, Shi Feng, Kaisong Song, Yifei Zhang, Ge Yu

Abstract—The sparseness and incompleteness of knowledge graphs (KGs) trigger considerable interest in enhancing the representation learning with external corpora. However, the difficulty of aligning entities and relations with external corpora leads to inferior performance improvement. Open knowledge graphs (OKGs) consist of entity-mentions and relation-mentions that are represented by noncanonicalized freeform phrases, which generally do not rely on the specification of ontology schema. The roughness of the nonontological construction method leads to a specific characteristic of OKGs: diversity, where multiple entity-mentions (or relation-mentions) have the same meaning but different expressions. The diversity of OKGs can provide potential textual and structural features for the representation learning of KGs. We speculate that leveraging OKGs to enhance the representation learning of KGs can be more effective than using pure text or pure structure corpora. In this paper, we propose a new **OERL**, **O**pen knowledge graph **E**nhanced **R**epresentation **L**earning of KGs. OERL automatically extracts textual and structural connections between KGs and OKGs, models and transfers refined profitable features to enhance the representation learning of KGs. The strong performance improvement and exhaustive experimental analysis prove the superiority of OERL over state-of-the-art baselines.

Index Terms—Representation Learning, Knowledge Graph Embedding, Open Knowledge Graph, Enhanced Representation Learning

1 INTRODUCTION

KNOWLEDGE graphs (KGs) aim to represent objective facts in structured forms that can visually express the potential connections of the facts reflected by entities and relations. From the perspective of simplicity and comprehensiveness, KGs are often stored in the form of triples, such as a triple (*head entity*, *relation*, *tail entity*), abbreviated as (*h*, *r*, *t*). To better explore rich and latent semantic information in KGs, much effort has been devoted to mapping entities and relations into vector representation spaces, which is often called representation learning or knowledge graph embedding [1], [2]. The learned representations make the knowledge graph essentially computable, which establishes the foundation for downstream tasks, such as link prediction [3] and node classification [4].

In the early stage of representation learning, many models focused on capturing the potential structural features of KGs. Typical translation-based representation learning models, such as TransE [1], TransH [5], TransD [6] and SimplE [7], employ transitional operations to model the relations between entities. DistMult [8] and ComplEx [9] use tri-linear operations to compute the similarity scores of triples in KGs. RotatE [10] concentrates on inferring various relation patterns by defining each relation as a rotation from the source entity to the target entity in a complex vector space. ConvE [2] applies a convolutional neural network to learn nonlinear complex features. The above traditional models have made great progress in representation learning of KGs. However, their performances are inevitably limited by the sparseness and incompleteness of KGs.

To address this problem, much work has proposed fusing external textual information to enhance the representation learning of KGs. External textual information refers to supplementary features, e.g., entity types, entity descriptions, relation features and world knowledge, which could be helpful in understanding KGs. SSE [11] and TKRL [12] take advantage of entity types to strengthen representation learning. DKRL [13] represents each entity with a structural vector and a descriptive vector to capture the textual information in entity descriptions. In addition to entity information, many models, such as TEKE [14] and AATE_E [15], incorporate supplementary features of relations into representation learning. Recent work [16], [17] employs pre-trained language models to enrich representation learning with world knowledge. In brief, text-enhanced representation learning models can capture both potential structural features in their own KGs and semantic properties in external supplementary corpora. However, external corpora are difficult to align with the entities and relations in KGs, which leads to insufficient performance improvements.

We have recently discovered that open knowledge graphs (OKGs) [18] are more general knowledge graphs that contain rich structural knowledge and textual information at the same time. OKGs [18], constructed based on open information extraction systems [19], regard noun phrases as entity-mentions and relation phrases as relation-mentions, and generally do not rely on the specification of ontology schema. The roughness of the nonontological construction method leads to a specific characteristic of OKGs – diversity: multiple entity-mentions (or relation-mentions) have the same meaning but different forms of expression. We speculate that this diversity of OKGs is conducive to enhancing the representation learning of KGs. For an entity in a KG, there could be multiple entity-mentions in an OKG which have the same meaning but different expressions as the entity. The same applies to relations. For example, in Figure 1(a), entity *NBC* in the KG has the same meaning as two entity-mentions *NBC Television* and *NBC-TV* in the OKG. Similarly, relation *has office in* in the KG has the same

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2022.3218850

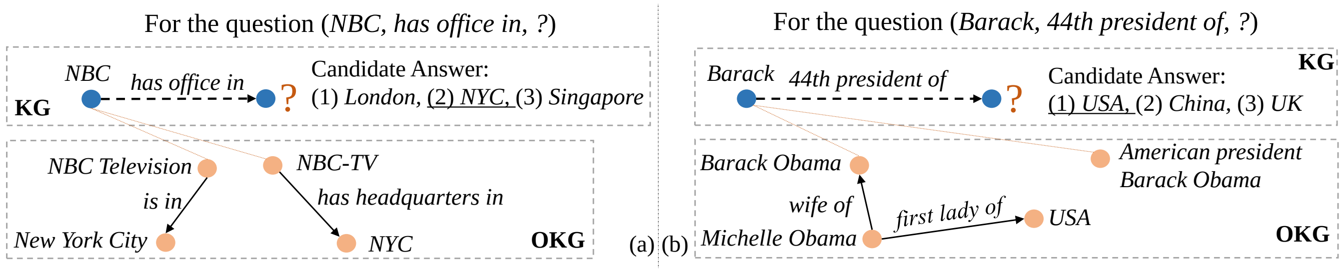


Fig. 1. Examples of OERL. (a) For the question (*NBC, has office in, ?*) in the KG, there is no evidence in the KG to prove its answer.. Entity-mentions *NBC Television*, *NBC-TV* in the OKG have the same meaning with the entity *NBC* in the KG. The related triples *(NBC Television, is in, New York City)* and *(NBC-TV, has headquarters in, NYC)* in the OKG can provide evidence to prove that the answer of (*NBC, has office in, ?*) is *NYC*. (b) For the question *(Barack, 44th president of,?)* in the KG, entity-mentions *Barack Obama*, *American president Barack Obama* and their structures in the OKG can provide more evidence to prove that the answer is entity *USA*.

meaning as two relation-mentions *is in* and *has headquarters in* in the OKG. These different expression forms of entity-mentions (relation-mentions) with the same meaning in the OKG can describe the entity (relation) in the KG from multiple perspectives, which provides rich textual information for the KG. Moreover, through the above connections of (entity, entity-mentions) and (relation, relation-mentions), we are able to extract profitable structures from the OKG to supplement the missing structural information of the KG. For example, in Figure 1(b), when the structural facts (*Michelle Obama, wife of, Barack Obama*) and (*Michelle Obama, first lady of, USA*) in the OKG are known, it can be inferred with great certainty that the answer to the link prediction question (*Barack, 44th president of, ?*) in the KG is *USA*. Based on the above observations, we infer that using OKGs to enhance representation learning of KGs is more effective than using pure text or pure structure corpora.

In this paper, we propose a new **OERL**, **O**pen knowledge graph **E**nhanced **R**epresentation **L**earning of knowledge graphs. OERL advocates enhancing representation learning of KGs with favorable potential information from OKGs. Specifically, OERL first extracts and models the *Textual Connection* ($C^{\mathcal{E}}$, $C^{\mathcal{R}}$) with textual rule extraction, textual semantic filtering and textual enhanced modeling modules, which can enhance KGs with textual features from OKGs. Then, OERL extracts and models the *Structural Connection* ($S^{\mathcal{E}}$) with structural ego extraction and structural enhanced modeling modules, which can provide rich structural semantics. Finally, OERL transfers refined profitable features from textual and structural connections to the representation learning of KGs. The strong performance improvement and exhaustive experimental analysis prove the superiority of OERL over state-of-the-art baselines. In summary, our key contributions are described as follows:

- We are the first to use the diversity properties of OKGs to enhance representation learning. This OKG enhanced strategy can be easily bootstrapped to new domains.
- We propose a new strategy, OERL, which automatically extracts textual and structural connections between KGs and OKGs, models and transfers refined profitable features to enhance the representation learning of KGs.
- We perform extensive experiments to show the superiority of OERL over state-of-the-art baselines. We have released our code and datasets at https://github.com/feiwangyuzhou/OERL.

The rest of this paper is structured as follows: We introduce related work in §2 and problem formulation in §3. Our OERL is proposed in §4. Experimental results and analysis are given in §5. Conclusion and future work are given in §6.

2 Related Work

2.1 Knowledge Graphs and Representation Learning

The purpose of KGs is to represent objective facts in a structured form, which can intuitively express the potential relation between facts reflected by entities and relations in KGs. Usually, KGs are stored in the form of triples, where these triples can not only store the characteristics of entities and relations but also save space in comparison with matrix storage. When using data for operations, this storage method can be easily transformed into a graph structure. KGs have been widely used in question answering [20], dialog systems [21] and recommender systems [22].

To better explore the latent semantic features of KGs, many efforts have been devoted to mapping entities and relations into vector representation space [3]. Translation-based representation learning models, such as TransE [1], TransH [5], TransR [23], and TransD [6], employ transitional characteristics to model entities and relations in KGs. DistMult [8] and ComplEx [9] use tri-linear operations to compute similarity scores for triples. SimplE [7] presents a simple enhancement of canonical polyadic decomposition, which allows two representations of each entity to be learned dependently. RotatE [10] defines each relation as a rotation from the source entity to the target entity in a complex vector space, and infers various relation patterns including: symmetry, antisymmetry, inversion, and composition. ConvE [2] applies a convolutional neural network to learn nonlinear features and capture complex relations in KGs. The above representation learning models can capture potential structural features of their own KGs. However, their performances are inevitably limited by the sparseness and incompleteness of KGs.

2.2 Text Enhanced Representation Learning

In view of the sparseness and incompleteness of KGs, fusion of external textual information in representation learning of KGs has been another important research direction [15]. Entity types are important features for the representation learning of KGs. Nickel et al. [24] transform entity types into an *isA* relation, establish triples (*head entity*, *isA*, *entity type*), and train these triples with the initial triples equally. SSE [11] takes full advantage of additional semantic information and ensures that the representation space is semantically smooth, which requires entities of the same type to remain close to each other in the representation space. TKRL [12] presents a type-embodied representation learning model to take advantage of hierarchical entity types.

Entity descriptions are considerable supplementary for the representation learning of KGs. NTN [25] learns word vectors from an auxiliary news corpus and initializes the representations of entities by averaging word vectors. Wang et al. [26] align the KG with an auxiliary textual corpus and combine the representation learning of KGs with that of words to better utilize external textual features. Long et al. [27] leverage entity descriptions in lexical resources in conjunction with distributional semantics to derive better initialization for representation learning. DKRL [13] represents each entity with a structural vector to capture the structural information in KGs and a description vector to capture the textual information in entity descriptions.

In addition to entities, many models incorporate relation features into representation learning of KGs. TEKE [14] expands the semantic structures of KGs with a textual corpus and gives each relation different representations for different head and tail entities, which can better handle 1-N, N-1 and N-N complex relations. AATE_E [15] proposes a mutual attention mechanism to exploit entity descriptions and triple specific relation-mentions, which gives each entity (relation) different representations in different triples. Recent work employs language models for representation learning of KGs. KG-Bert [16] encodes textual descriptions of entities and relations with a bidirectional transformer mechanism. Pretrain-KGE [17] enriches the representation learning of KGs via pretrained language models and proposes a universal training framework with a semantic-based fine-tuning phase, knowledge extracting phase and KGE training phase.

In brief, text-enhanced representation learning can capture both the structural features in their own KGs, and the semantic properties in external supplementary corpora. However, the external supplementary corpora are difficult to align with the entities and relations in KGs, which leads to insufficient performance improvements.

2.3 Open Knowledge Graphs and Related Models

Open information extraction systems [28], such as ReVerb [29], OLLIE [30], ClauseIE [31], RelNoun [32], BONIE [33], CALMIE [34] and OPIEC [19], automatically extract triples (subject noun phrase, relation phrase, object noun phrase) from raw data. OKGs are constructed based on open information extraction systems, regarding subject and object noun phrases as entity-mentions and relation phrases as relation-mentions. The entity-mentions (relation-mentions) of OKGs are represented by *noncanonicalized*, *free-form phrases*, and generally do not rely on the specification of ontology schema. OpenKGs tend to be nosier without specification of ontology schema, so much work performs automatic canonicalization to canonicalize the naming of entities and relations. Logical rules and cluster algorithms are adopted to classify relation phrases [35] and noun phrases [36]. CESI [37] performs canonicalization with hierarchical agglomerative clustering over learned embeddings of OKGs by combining embeddings and relevant phrase side information. MGNN [38] proposes a graph neural network model to canonicalize OKGs by aggregating the representations of noun phrases and relation phrases through a multi-layered metagraph structure. However, the roughness of the nonontological construction method also leads to a specific characteristic of OKGs – diversity: multiple entity-mentions (or relation-mentions) have the same meaning but different expression forms. For an entity (relation) in a KG, there may be multiple entity-mentions (relation-mentions) in an OKG, that have the same meaning but different expression forms as the entity (relation). The connections between OKGs and KGs can provide rich textual and structural information for the representation learning of KGs. Therefore, we propose using OKGs to enhance the representation learning of KGs.

3 Problem Formulation

Definition 1. Knowledge Graph (KG) and Representations In a KG $\mathcal{G}$=$(\mathcal{E}, \mathcal{R})$, let a triple be (h, r, t), where $h, t \in \mathcal{E}$ represent entities and $r \in \mathcal{R}$ represents the relation between entities h and t. In the KG, the meanings of any entity (relation) pairs are completely different [39]. The aim of representation learning is to learn the representations of entities and relations that carry as much key information as possible. The initial entity representations and relation representations are $\mathbf{E} \in \mathbb{R}^{N_e \times T}$ and $\mathbf{R} \in \mathbb{R}^{N_r \times T}$, respectively, where N_e, N_r are the numbers of entities and relations and T is the feature dimension. $\mathbf{h} \in \mathbf{E}$ is the initial representation of entity h, and $\mathbf{r} \in \mathbf{R}$ is the initial representation of relation r.

Definition 2. Open Knowledge Graph (OKG) and Representations In an OKG $\mathcal{G}^o$=$(\mathcal{E}^o, \mathcal{R}^o)$, let a triple be (h^o, r^o, t^o), where $h^o, t^o \in \mathcal{E}^o$ represent entity-mentions and $r^o \in \mathcal{R}^o$ represents the relation-mention between entity-mentions h^o and t^o. In the OKG, there are multiple entity-mentions (relation-mentions) with the same meaning but different expression forms [18]. Entity-mentions h^o, t^o and relation-mention r^o are represented by nonempty sequences of words from a word vocabulary, where w_{h^o}=$\{w_{h^o,i}\}_{i=1}^{|w_{h^o}|}$, w_{r^o}=$\{w_{r^o,i}\}_{i=1}^{|w_{r^o}|}$ are the word sequence of entity-mention h^o and relation-mention r^o, respectively. $|w_{h^o}|, |w_{r^o}|$ are the numbers of words.

Definition 3. Open Knowledge Graph Enhanced Representation Learning We propose a new strategy, OERL, to enhance the representation learning of KGs with potential features of OKGs. We now define the salient concepts of connections that underlie the enhanced strategy. The enhanced connections include two parts: textual connections $\{C^{\mathcal{E}}, C^{\mathcal{R}}\}$ and structural connection $S^{\mathcal{E}}$. For the textual connection, $C^{\mathcal{E}}$=$\{C_h^{\mathcal{E}}\}_{h \in \mathcal{E}}$ is the set for entities. Specifically, for an entity h in the KG, $C_h^{\mathcal{E}}$=$\{h_i^o\}_{i=1}^{|C_h^{\mathcal{E}}|}$ is the set of related entity-mentions in the OKG. Similarly, $C^{\mathcal{R}}$=$\{C_r^{\mathcal{R}}\}_{r \in \mathcal{R}}$ is the set for relations, and $C_r^{\mathcal{R}}$=$\{r_i^o\}_{i=1}^{|C_r^{\mathcal{R}}|}$ is the set of related relation-mentions in the OKG for relation r in the KG. $|C_h^{\mathcal{E}}|$, $|C_r^{\mathcal{R}}|$ are the number of mentions in the set $C_h^{\mathcal{E}}$, $C_r^{\mathcal{R}}$. For the structural connection, $S^{\mathcal{E}}$=$\{S_h^{\mathcal{E}}\}_{h \in \mathcal{E}}$ is the structual connection for all entities in the KG. Concretely, for entity h in the KG, its structural connection is $S_h^{\mathcal{E}}$=$\{S(h, h_i^o)\}_{h_i^o \in C_h^{\mathcal{E}}}$,

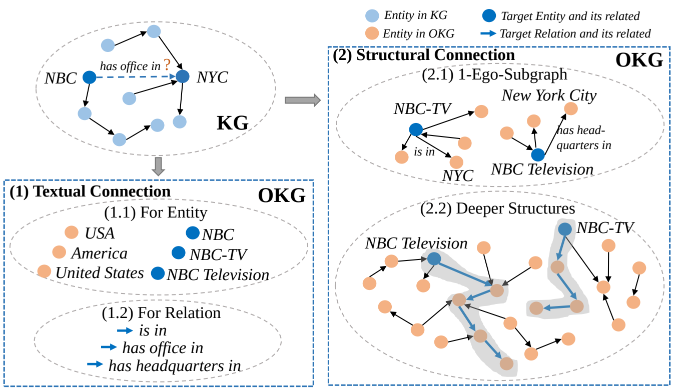


Fig. 2. Examples of enhanced types from OKGs. (1) Textual Connection: entity-mentions and relation-mentions in the OKG provide a variety of description information for related entities and relations in the KG. (2) Structural Connection: 1-ego-subgraphs and deeper structures in the OKG give structural features to enhance the KG.

in which $S(h, h_i^o)=\{\{(h_i^o, r_j^o, t_j^o)\}_{j=1}^{N_i^{\rightarrow}}, \{(t_j^o, r_j^o, h_i^o)\}_{j=1}^{N_i^{\leftarrow}}\}$ is the related subgraph of entity-mention h_i^o in the OKG.

4 Proposed OERL

In this section, we propose a new OERL strategy to enhance representation learning of KGs with potential features from OKGs. Figure 3 presents the overall framework of OERL. In §4.1, we extract and model the *Textual Connection* ($C^{\mathcal{E}}$, $C^{\mathcal{R}}$) with textual rule extraction, textual semantic filtering and textual enhanced modeling, which can enhance KGs with textual features from OKGs. In §4.2, we extract and model the *Structural Connection* ($S^{\mathcal{E}}$) with structural ego extraction and structural enhanced modeling, which can provide rich structural semantics for representation learning of KGs. In §4.3, we transfer refined profitable features from textual and structural connections to the representation learning of KGs.

4.1 Textual Connection

We extract textual connections ($C^{\mathcal{E}}$, $C^{\mathcal{R}}$) between the KG and the OKG, where $C^{\mathcal{E}}$ is the set of textual connections between entities (in the KG) and entity-mentions (in the OKG), $C^{\mathcal{R}}$ is the set of textual connections between relations (in the KG) and relation-mentions (in the OKG). For an entity h in the KG, multiple entity-mentions in its textual connection $C_h^{\mathcal{E}}$, can supply various sufficient descriptions for the entity h, which could be beneficial for enhancing the representation learning of KGs. The same is true of relations. For example, in Figure 1(b), entity *Barack* in the KG has the same meaning as entity-mentions *Barack Obama* and *American president Barack Obama* in the OKG. For the link prediction question (*Barack, 44th president of, ?*), there may be no evidence in the KG to prove that *(Barack, is the 44th president of, USA)*, while the entity-mention *American president Barack Obama* in the OKG can provide helpful information. Therefore, we believe that OKGs can supplement the missing textual information of KGs, which is the motivation to build textual connections. In the following, $C^{\mathcal{E}}$ and $C^{\mathcal{R}}$ are extracted with rule extraction and semantic filtering, then modeled with enhanced modeling module.

Textual Rule Extraction (TRE) For each entity (relation) in the KG, the set of related entity-mentions (relation-mentions) in the OKG are extracted with manually designed rules. This set should cover relevant entity-mentions (relation-mentions) as widely and comprehensively as possible. The extraction rules for an entity h in the KG are as follows. (1) Synonym Searching: For h, we search for synonyms with the Wordnet tool [40] and put these synonyms to set $Syn(h)$. This operation can expand the search scope and find as many entity-mentions with the same semantics as possible. (2) Fuzzy Matching [1]: For each synonym in $Syn(h)$, we extract related entity-mentions from the OKG with the IDF-Token-Overlap tool [36]. The related entity-mentions of all synonyms are extracted with Step (2) and merged to set $C_h^{\mathcal{E}}$. An analogous description follows for the relation r in the KG to extract the set of related relation-mentions $C_r^{\mathcal{R}}$. However, textual connections ($C_h^{\mathcal{E}}$, $C_r^{\mathcal{R}}$) based on the above automatic rules are large and noisy. To reduce these large and noisy sets, we further extract meaningful information from the following semantic similarity.

Textual Semantic Filtering (TSF) We reduce the noise of $C_h^{\mathcal{E}}$, $C_r^{\mathcal{R}}$ according to semantic similarity. Next, we take $C_h^{\mathcal{E}}$ as an example to explain in detail. For entity h and its textual connection $C_h^{\mathcal{E}}$, we compute the semantic similarity scores between entity h and each entity-mention $h_i^o \in C_h^{\mathcal{E}}$, then remove the mentions whose scores are below a threshold. The similarity score of any pair is computed with a cosine similarity function as:

$$\beta(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}^\top \mathbf{b}}{||\mathbf{a}|| \cdot ||\mathbf{b}||} \tag{1}$$

where $\top$ indicates the transpose. The similarity score between entity h and entity-mention h_i^o is denoted as $\beta(\widetilde{\mathbf{h}}, \widetilde{\mathbf{h_i^o}})$:

$$\widetilde{\mathbf{h}} = f(w_h), \quad \widetilde{\mathbf{h_i^o}} = f(w_{h_i}^o) \tag{2}$$

in which $w_h, w_{h_i^o}$ are word sequences of entity h and entity-mention h_i^o, respectively. f represents the method to learn semantic features that we used here is BERT [2] [41]. The format of input to BERT is [CLS] + w_h ($w_{h_i^o}$) + [SEP] for entity h (entity-mention h_i^o) and [CLS] + w_r ($w_{r_i^o}$) + [SEP] for relation r (relation-mention r_i^o). We use representations of [CLS] token from the last layer of BERT as final representations. Through this semantic filtering, mentions in $C_h^{\mathcal{E}}$, $C_r^{\mathcal{R}}$ with irrelevant semantics can be removed in large numbers.

Thus, textual connections $C_h^{\mathcal{E}}$ for entity h and $C_r^{\mathcal{R}}$ for relation r have been built. These textual connections can provide a variety of description information to enhance the representation learning of entities and relations in the KG (Fig. 2(1)). Concurrently, instead of assigning each entity-mention (relation-mention) to exactly one entity (relation), the above extraction mechanism is soft, where an entity-mention (relation-mention) with different senses (polysemy) can be assigned to different entities (relations). According to our statistics, most entities and relations in KG can search for relevant mentions in the OKG, and the number of mentions

1. In fuzzy matching, tokens in both synonyms and entity-mentions are restored with the NLTK part-of-speech tagging tool, which can reduce the omission caused by different parts of speech.

2. This part is flexible for other pretrained sequence algorithms. BERT we used here is the base version and its parameters are fixed.

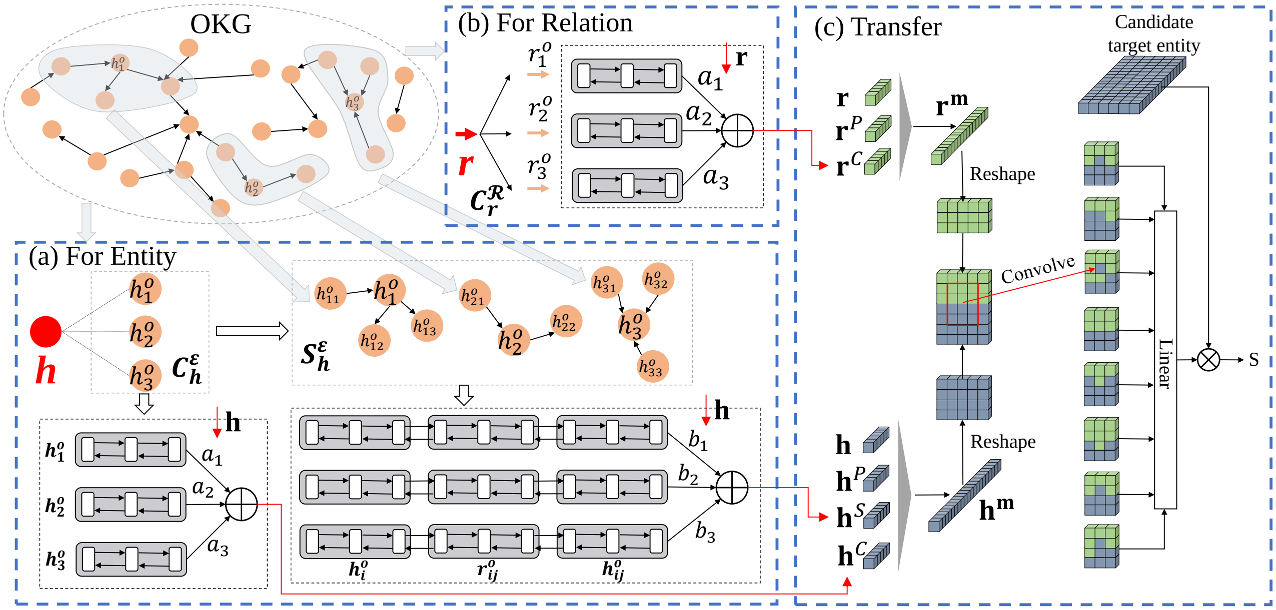


Fig. 3. Framework of the proposed OERL. (a) For entity h in the KG, we extract textual connection $C_h^{\mathcal{E}}$ and structural connection $S_h^{\mathcal{E}}$ from the OKG, and use the attention mechanism to refine profitable features from both connections. (b) For relation r in the KG, we extract and model textual connection $C_r^{\mathcal{R}}$. (c) The above refined profitable features of entity h and relation r are transferred to the KG.

is often greater than 1, that is, $|C_h^{\mathcal{E}}|>1$, $|C_r^{\mathcal{R}}|>1$. However, whether in $C_h^{\mathcal{E}}$ or $C_r^{\mathcal{R}}$, some mentions are beneficial to the entity h or relation r, while other mentions are meaningless or noisy. Next, we use the enhanced modeling module to further absorb favorable information from $C_h^{\mathcal{E}}$ and $C_r^{\mathcal{R}}$.

Textual Enhanced Modeling (TEM) Here, we explore how to aggregate the connection features from the above textual connections by discarding noise data and absorbing favorable information. Multiple entity-mentions in $C_h^{\mathcal{E}}$ and relation-mentions in $C_r^{\mathcal{R}}$ can supply various description information, which could be beneficial for enhancing the representation learning of entity h and relation r. However, the mentions in $C_h^{\mathcal{E}}$ and $C_r^{\mathcal{R}}$ encounter two phenomena: (1) Noise: Some mentions are unrelated, and (2) Importance: The importance of different useful (not noisy) mentions is different, where some mentions can introduce more external features than others. Therefore, it is necessary to design an effective method to discard noise data and absorb favorable information. We introduce an attention mechanism to assign different weights to different data according to the needs of the situation. We hope that the favorable information can be assigned a large weight, and the noisy data is assigned a little weight. The enhanced information in $C_h^{\mathcal{E}}$ or $C_r^{\mathcal{R}}$ can be aggregated with the weights. Therefore, the textual enhanced embeddings for entity h ($\mathbf{h}^C$) and for relation r ($\mathbf{r}^C$) are computed separately as:

$$\mathbf{h}^C = \sum_{h_i^o \in C_h^{\mathcal{E}}} a_i^h \mathbf{h_i^o}, \quad C_h^{\mathcal{E}} = \{h_i^o\}_{i=1}^{|C_h^{\mathcal{E}}|} \tag{3}$$

$$\mathbf{r}^C = \sum_{r_i^o \in C_r^{\mathcal{R}}} a_i^r \mathbf{r_i^o}, \quad C_r^{\mathcal{R}} = \{r_i^o\}_{i=1}^{|C_r^{\mathcal{R}}|} \tag{4}$$

$\mathbf{h_i^o}$ and $\mathbf{r_i^o}$ are the representation of entity-mention h_i^o and relation-mention r_i^o, in which we use textual technology (Enc), such as BiGRU [42] and BERT [41], to learn textual features of entity-mention h_i^o and relation-mention r_i^o:

$$\mathbf{h_i^o} = \mathrm{Enc}(w_{h_i^o}), \quad \mathbf{r_i^o} = \mathrm{Enc}(w_{r_i^o}) \tag{5}$$

where $w_{h_i^o}=\{w_{h_i^o,j}\}_{j=1}^{|w_{h_i^o}|}$, $w_{r_i^o}=\{w_{r_i^o,j}\}_{j=1}^{|w_{r_i^o}|}$ are the word sequences of entity-mention h_i^o and relation-mention r_i^o. And a_i^h in Eq. (3) is the attention weight of entity-mention h_i^o to entity h, and a_i^r in Eq. (4) is the attention weight of relation-mention r_i^o to relation r. The attention weights can be computed as:

$$a_i^h = \frac{\exp\big(\mathrm{LeakyR}(\beta(\mathbf{h}, \mathbf{h_i^o}))\big)}{\Sigma_{h_k^o \in C_h^{\mathcal{E}}} \exp\big(\mathrm{LeakyR}(\beta(\mathbf{h}, \mathbf{h_k^o}))\big)} \tag{6}$$

$$a_i^r = \frac{\exp\big(\mathrm{LeakyR}(\beta(\mathbf{r}, \mathbf{r_i^o}))\big)}{\Sigma_{r_k^o \in C_r^{\mathcal{R}}} \exp\big(\mathrm{LeakyR}(\beta(\mathbf{r}, \mathbf{r_k^o}))\big)} \tag{7}$$

in which LeakyR is the LeakyRelu nonlinearity function $\beta(a, b)$ is the cosine similarity function in Eq. (1). As is known to all, each entity and relation in the KG could be assigned a unique initialized and trainable embedding, where $\mathbf{h} \in \mathbf{E}, \mathbf{r} \in \mathbf{R}$ are the trainable embedding for entity h and relation r, respectively.

To date, textual enhanced embeddings for entity h ($\mathbf{h}^C \in \mathbb{R}^T$) and for relation r ($\mathbf{r}^C \in \mathbb{R}^T$) have been captured, which will be transferred to the KG in §4.3.

4.2 Structural Connection

Taking textual connections $\{C^{\mathcal{E}}, C^{\mathcal{R}}\}$ as the base, we construct the structural connection $S^{\mathcal{E}}$ between the KG and the OKG to enhance the representation learning from the perspective of structure. Structural connection $S^{\mathcal{E}}$ is designed only for entities and not for relations in the KG [3]. More precisely, structural connection $S^{\mathcal{E}}$ uses structural features in

3. To our knowledge, only entities have structural features in the KG, and relations are used to assist entities in forming connections.

the OKG to enhance the representation learning of entities in the KG. Next, $S^{\mathcal{E}}$ is extracted with structural ego extraction, then modeled with structural enhanced modeling.

Definition 4: An *m*-ego subgraph of an entity is a subgraph in which the entity can connect with incoming and outgoing paths whose hop $\leq$ m.

Structural Ego Extraction (SEE) We propose using subgraphs as structural features by extending each textual connection to its local structure. Specifically, for a certain entity h and a textual connection $h_i^o \in C_h^{\mathcal{E}}$, we define the structural connection $S(h, h_i^o)$ to be the 1-ego subgraph of entitiy-mention h_i^o-[4]. According to the definition of the *m*-ego subgraph, the 1-ego subgraph of entity-mention h_i^o is a subgraph in which the entity-mention h_i^o can connect with directed links (paths whose hop $\leq$ 1), including both incoming and outgoing directions. We extract and store the 1-ego subgraph of entity-mention h_i^o into $S(h, \underline{h_i^o})$=$\{\{(\underline{h_i^o}, r_j^o, t_j^o)\}_{j=1}^{N_{out}}, \{(t_j^o, r_j^o, \underline{h_i^o})\}_{j=1}^{N_{in}}\}$, and place these multiple 1-ego subgraphs of all textual connections into the set $S_h^{\mathcal{E}} = \{S(h, h_i^o)\}_{h_i^o \in C_h^{\mathcal{E}}}$. Thus far, structural connection $S_h^{\mathcal{E}}$ has been completed. The structural connection can supply rich structural features for representation learning of KGs. For example, in Fig. 2(2.1), the 1-ego subgraph of entity-mention NBC-TV is extracted from the OKG, with one incoming link and three outgoing links. Among them, the outgoing link (*NBC-TV, is in, NYC*) supplies missing and pivotal evidence for predicting the link (*NBC, has office in, NYC*) in the KG. Through $S^{\mathcal{E}}$, we can further mine deeper structural features from OKGs to enhance the representation learning of KGs (Fig. 2(2.2)), which could be future work. Next, we use structural enhanced modeling to further absorb favorable information from $S^{\mathcal{E}}$.

Structural Enhanced Modeling (SEM) Here, we explore how to aggregate the connection features from the structural connection. For an entity h in the KG, its structural connection $S_h^{\mathcal{E}} = \{S(h, h_i^o)\}_{h_i^o \in C_h^{\mathcal{E}}}$ is able to provide structural semantic understanding, which could be beneficial for enhancing the representation learning of entity h. However, these structural features are not equally important, and only part of the structure can promote the current entity h. Therefore, designing an effective method to identify useful structural information in $S_h^{\mathcal{E}}$ is an important and indispensable step. as mentioned in §4.1, we have designed an attention mechanism to assign different weights to different entity-mentions in $C_h^{\mathcal{E}}$, which is based on the text description of entity-mentions. Here, these are structural features contained in the triples, and the triples have two directions for entity-mention h_i^o: outgoing $\{(\underline{h_i^o}, r_j^o, t_j^o)\}_{j=1}^{N_i^{\rightarrow}}$ and incoming $\{(t_j^o, r_j^o, \underline{h_i^o})\}_{j=1}^{N_i^{\leftarrow}}$. We use textual technology (Enc which is similar to Eq. (5)) to model the directed triples:

$$[\mathbf{h_i^o}; \mathbf{r_j^o}; \mathbf{t_j^o}] = \text{Enc}(w_{h_i^o}; w_{r_j^o}; w_{t_j^o}),\ \text{for}\{(\underline{h_i^o}, r_j^o, t_j^o)\}_{j=1}^{N_i^{\rightarrow}} \quad (8)$$

$$[\mathbf{t_j^o}; \mathbf{r_j^o}; \mathbf{h_i^o}] = \text{Enc}(w_{t_j^o}; w_{r_j^o}; w_{h_i^o}),\ \text{for}\{(t_j^o, r_j^o, \underline{h_i^o})\}_{j=1}^{N_i^{\leftarrow}} \quad (9)$$

where $[\mathbf{h_i^o}; \mathbf{r_j^o}; \mathbf{t_j^o}]$ and $[\mathbf{t_j^o}; \mathbf{r_j^o}; \mathbf{h_i^o}]$ are the representations of the outgoing triple (h_i^o, r_j^o, t_j^o) and incoming triple (t_j^o, r_j^o, h_i^o), respectively. $(w_{h_i^o}; w_{r_j^o}; w_{t_j^o})$ represents the concatenation of the word sequences of entity-mention h_i^o, relation-mention r_j^o and entity-mention t_j^o in order, as well as $(w_{t_j^o}; w_{r_j^o}; w_{h_i^o})$. Then, we use the attention mechanism to assign different weights to different triples:

$$a_{ij}^{\rightarrow} = \frac{\exp(\text{LeakyR}(\beta(\mathbf{h}, [\mathbf{h_i^o}; \mathbf{r_j^o}; \mathbf{t_j^o}])))}{\Sigma_{(h_k^o, r_k^o, t_k^o) \in S_h^{\mathcal{E}}} \exp(\text{LeakyR}(\beta(\mathbf{h}, [\mathbf{h_k^o}; \mathbf{r_k^o}; \mathbf{t_k^o}])))} \quad (10)$$

$$a_{ij}^{\leftarrow} = \frac{\exp(\text{LeakyR}(\beta(\mathbf{h}, [\mathbf{t_j^o}; \mathbf{r_j^o}; \mathbf{h_i^o}])))}{\Sigma_{(h_k^o, r_k^o, t_k^o) \in S_h^{\mathcal{E}}} \exp(\text{LeakyR}(\beta(\mathbf{h}, [\mathbf{h_k^o}; \mathbf{r_k^o}; \mathbf{t_k^o}])))} \quad (11)$$

where $a_{ij}^{\rightarrow}, a_{ij}^{\leftarrow}$ are the weights of the outgoing triple (h_i^o, r_j^o, t_j^o) and incoming triple (t_j^o, r_j^o, h_i^o). Ultimately, enhanced information in $S_h^{\mathcal{E}}$ can be aggregated with the weights:

$$\mathbf{h}^S = \sum_{h_i^o \in S_h^{\mathcal{E}}} \left(\Sigma_{j=1}^{N_i^{\rightarrow}} a_{ij}^{\rightarrow} [\mathbf{h_i^o}; \mathbf{r_j^o}; \mathbf{t_j^o}] + \Sigma_{j=1}^{N_i^{\leftarrow}} a_{ij}^{\leftarrow} [\mathbf{t_j^o}; \mathbf{r_j^o}; \mathbf{h_i^o}] \right) \quad (12)$$

Structural enhanced representation $\mathbf{h}^S \in \mathbb{R}^T$ of entity h has been aggregated and will be transferred to the KG in §4.3.

4.3 Transfer to KG

This section focuses on transferring the enhanced representations to the representation learning of the KG and helping the KG achieve better performance in downstream tasks. Through the above module, we have obtained the enhanced representation $\mathbf{h}^C$, $\mathbf{h}^S$ of entity h and $\mathbf{r}^C$ for relation r. Next, the enhanced representations are merged with the original representation in the KG.

$$\begin{aligned} \mathbf{h^m} &= \mathbf{h} \oplus \mathbf{h}^P \oplus \mathbf{h}^C \oplus \mathbf{h}^S \\ \mathbf{r^m} &= \mathbf{r} \oplus \mathbf{r}^P \oplus \mathbf{r}^C \end{aligned} \quad (13)$$

where $\mathbf{h}^P = \text{Enc}(w_h)$, $\mathbf{r}^P = \text{Enc}(w_r)$ are the textual features of entity h and relation r in the KG. After concatenation $\oplus$, we obtain merging entity embedding $\mathbf{h^m} \in \mathbb{R}^{4T}$ and merging relation embedding $\mathbf{r^m} \in \mathbb{R}^{3T}$.

Then, we focus on exploiting potential connections between entities and relations. We use a two-dimensional convolutional network [2] to learn potential connections between a head entity h and a relation r as follows:

$$\varphi(h, r) = \sigma\big(\text{Linear}(\sigma(\text{Conv2d}_{\omega}([\widehat{\mathbf{h^m}}; \widehat{\mathbf{r^m}}])))\big) \quad (14)$$

where σ represents a ReLU activation, and $\widehat{\mathbf{h^m}}$ and $\widehat{\mathbf{r^m}}$ denote a reshaping of $\mathbf{h^m}$ and $\mathbf{r^m}$, respectively. Specifically, the reshaping operation converts the merging entity embedding $\mathbf{h^m} \in \mathbb{R}^{4T}$ from one dimension to two dimensions $\widehat{\mathbf{h^m}} \in \mathbb{R}^{T_{h1} \times T_{h2}}$, where $4T = T_{h1}T_{h2}$. The reshaping operation converts the merging relation embedding $\mathbf{r^m} \in \mathbb{R}^{3T}$ from one dimension to two dimensions $\widehat{\mathbf{r}} \in \mathbb{R}^{T_{r1} \times T_{r2}}$, where $3T = T_{r1}T_{r2}$. Here, let $T_{h2} = T_{r2}$. $[\widehat{\mathbf{h^m}}; \widehat{\mathbf{r^m}}] \in \mathbb{R}^{(T_{h1}+T_{r1}) \times T_{h2}}$ represents the concatenation of the reshaped embeddings of $\widehat{\mathbf{h^m}}$ and $\widehat{\mathbf{r^m}}$ in the first dimension. The Conv2d$_\omega$ symbol denotes a two-dimensional convolutional layer with filters ω. This layer returns a feature map tensor $\mathcal{F} \in \mathbb{R}^{F_1 \times F_2 \times F_3}$, where F_1 is the number of feature maps of dimensions $F_2 \times F_3$. $\mathcal{F}$ is then reshaped into $\mathbb{R}^{F_1 F_2 F_3}$ and projected to $\mathbb{R}^T$ with a Linear layer. Through the convolution module, potential embeddings of entity h and relation r are jointly encapsulated.

The representation learning of KGs in this paper takes link prediction task as the downstream task. The aim of the

4. Expanding the scope from the 1-ego subgraph to *m*-ego subgraph ($m > 1$) worsen the performance.

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2022.3218850

link prediction task is to predict whether a triple is correct or incorrect. The similarity score for a triple (h, r, t) is:

$$S(h, r, t) = \varphi(h, r)^{\top}\mathbf{t} \tag{15}$$

where $\mathbf{t}$ is the representation of entity t.

For training, we use a binary cross-entropy loss to optimize parameters, where the correct samples are considered positive instances $\varrho^{+} = \{(h, r, t_j^{+})\}_{j=1}^{|\varrho^{+}|}$, and negative instances $\varrho^{-} = \{(h, r, t_j^{-})\}_{j=1}^{|\varrho^{-}|}$ are generated by corrupting the true tail entities, where a negative entity t_j^{-} is selected randomly from an entity list defined by: $\mathcal{E} - \mathcal{E}(h, r)$ with $\mathcal{E}(h, r)$ being the entity list of true answers (tail entities), that is, $t_i \in \mathcal{E}(h, r)$ if the triple $(h, r, t_i) \in \varrho^{+}$. The loss function of the proposed model is:

$$L = -\sum_{i=1}^{|\varrho^{+}|+|\varrho^{-}|} \frac{(Y_i \cdot \log X_i + (1 - Y_i) \cdot \log(1 - X_i))}{|\varrho^{+}| + |\varrho^{-}|} \tag{16}$$

$$X_i = \sigma(S(h_i, r_i, t_i)) \tag{17}$$

$$\begin{cases} Y_i = 1, if\ (h_i, r_i, t_i) \in \varrho^{+} \\ Y_i = 0, if\ (h_i, r_i, t_i) \in \varrho^{-} \end{cases} \tag{18}$$

When predicting the head h based on pair (r, t), we reverse the relation by adding a special symbol to obtain a new triple (t, r_{rev}, h) and then train it with Eq. (16)-Eq. (18).

5 Experiment

5.1 Datasets

Our motivation in this paper is to extract useful information from the OKG to enhance the representation learning of the KG. The datasets used here include an OKG and two KGs. Next, we first introduce the datasets in detail and then provide the connection statistics between the OKG and two KGs at the end. The details of all KGs are shown in Table 1.

- **An OKG**

OLPBENCH [18] is a recently published OKG dataset that regards noun phrases as entity-mentions and relation phrases as relation-mentions, which was extracted from English Wikipedia with the open information extraction system OPIEC [19]. In *OLPBENCH*, there are many entity-mentions with the same meaning but different expressions, e.g., *NBC-TV*, *NBC Television* and *NBC*, as well as relation-mentions, e.g., *has headquarters in* and *has office in*. This kind of synonymy in the OKG can help related entities (relations) in the KG to understand semantics from multiple perspectives. *OLPBENCH* also carries a large amount of noise data that should be limited as much as possible. In particular, entity-mentions contain two kinds of intolerable noise data: (1) entity-mentions that are pure numbers, which can be deleted by the type of conversion operations, (2) entity-mentions that are not noun phrases, where we use NLTK part-of-speech tagging tools to remove entity-mentions composed of adjective phrases or verb phrases to ensure that entity-mentions are noun phrases. Finally, this *OLPBENCH* with 360.8k entity-mentions, 286.0k relation-mentions and 20408.5k triples is used as the OKG to supply external information for KGs.

- **Two KGs**

(a) entity cloud (b) relation cloud

Fig. 4. Cloud charts of entities and relations on *OEKG* dataset.

CN100K [43] is a common sense KG that contains general common sense facts. The most important reason why we chose this *CN100K* as a benchmark KG is that the entities of *CN100K* are free-form text: the corresponding information is likely to be found from the OKG whose entities are noun phrases. This *CN100K* contains 78.2k entities, 34 relations, 100k train triples, 1.2k valid triples and 1.2k test triples.

OEKG is a subgraph extracted from the above OKG. We first randomly extract a subgraph from the above OKG and then remove synonymous entity-mentions (relation-mentions) to construct a new KG. For the removal step, we use the Textual Rule Extraction and Textual Semantic Filtering modules in §4.1, to cluster the entity-mentions and relation-mentions, where entity-mentions (relation-mentions) with similar meanings are grouped into one cluster. For each cluster, only one mention is retained, all others are deleted, and related triples of the deleted mentions are also deleted from the extracted subgraph. This difference in meaning between clusters satisfies the distinguishing characteristic of a KG: the meanings of any entity (relation) pairs are completely different. The constructed OEKG contains 105.9k entities, 34.6k relations, 250.1k train triples, 10.9k valid triples and 11.0k test triples. Figure 4 displays the cloud charts of entities and relations according to their frequencies, which can provide a macroscopic understanding of OEKG.

- **Connections**

Table 1 shows the details of textual and structural connections, where $|C^{\mathcal{R}}|_{\geq 1}$, $|C^{\mathcal{E}}|_{\geq 1}$ represent the number of entities and relations in the KG which has at least one entity-mention (relation-mention) in the OKG, $|S^{\mathcal{E}}|_{\geq 1}$ represents the number of entities in the KG which has at least one related triple in the OKG, $|C^{\mathcal{R}}|_{Avg}$, $|C^{\mathcal{E}}|_{Avg}$, $|S^{\mathcal{E}}|_{Avg}$ represent the average number of relation-mentions per relation, entity-mentions per entity, and related triples per entity. From the statistics in Table 1, 88% of relations for all datasets can find their related relation-mentions from the OKG, and 90% of entities for all datasets have related entity-mentions and triples from the OKG. The average number of relation-mentions per relation is greater than 25, and that of related triples per entity is greater than 30. The average number of entity-mentions per entity is 27.1 in *CN100K* and 7.4 in *OEKG*. The textual information in entity-mentions (relation-mentions) and structural features in triples is helpful to enhance the corresponding entity (relation).

5.2 Compared Models

Since our OKG enhanced strategy is proposed for the first time, resulting in no corresponding baseline that uses in-

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2022.3218850

TABLE 1
Details of CN100K, WN18RR and OEKG datasets.

Datasets	Ent	Rel	Train	Valid	Test	Density	$\|C^{\mathcal{R}}\|_{\geq 1}$	$\|C^{\mathcal{E}}\|_{\geq 1}$	$\|S^{\mathcal{E}}\|_{\geq 1}$	$\|C^{\mathcal{R}}\|_{Avg}$	$\|C^{\mathcal{E}}\|_{Avg}$	$\|S^{\mathcal{E}}\|_{Avg}$
CN100K	78.3k	34	100k	1.2k	1.2k	1.6e-5	30 / 88%	73.5k / 94%	72.6k / 93%	29.9	27.1	31.8
OEKG	39.6k	4.8k	68.3k	2.8k	2.9k	4.4e-5	4.4k / 92%	35.9k / 91%	35.7k / 90%	25.3	7.4	30.3

formation from OKGs, we propose multiple OERL variants for comparison with our approach. The compared models used in this paper include the following aspects: General Baselines are state-of-the-art representation learning models that do not use any external information, Model Variants are variants of standard OERL which use different models, and Data Variants are variants of standard OERL which use different types of enhanced data.

- **General Baselines**
  TransE [1] is a translation-based model to capture the entity and relation embeddings based on $h_k + s_m \approx h_j$.
  TransH [5] models a relation as a hyperplane together with a translation operation to well preserve the mapping properties of relations: reflexive, 1-N, N-1 and N-N.
  TransD [6] uses two vectors to represent a named symbol object (entity and relation), where the first vector represents the meaning of an entity (relation), and the second vector is used to dynamically construct a mapping matrix.
  DistMult [8] is a diagonal model that uses weighted elementwise dot products to learn entity embeddings and relation embeddings.
  ComplEx [9] is a complex valued embedding model that can handle a large variety of binary relations, including symmetric and antisymmetric relations.
  SimplE [7] is a tensor factorization approach based on canonical polyadic, that addresses the independence among the two embedding vectors of entities.
  RotatE [10] defines each relation as a rotation from the source entity to the target entity in the complex vector space, which infers various relation patterns including symmetry, antisymmetry, inversion, and composition.
  ConvE [2] uses 2D convolutional operations to capture global nonlinear features for entity embeddings and relation embeddings.
  KBGAT [44] introduces graph attention networks (GATs) [45] to incorporate entity and relation features in the proximity of each entity.
  GGAE [3] propose a global graph attention embedding network by combining global information from both direct neighbors and multihop neighbors.

- **Model Variants** Here, we provide some variant baselines of the proposed OERL model.
  OERL+Basic is the basic version of the OERL model, which uses the text descriptions of entities and relations contained in the KG, but does not use the textual connections and structural connections from OKGs and does not use the attention mechanism.
  OERL+MLP uses the two-layer MLP method instead of the attention mechanism in Eqs. (6), (7), (10) and (11), where the enhanced modeling methods are changed while the enhanced data remains unchanged.
  OERL+Mean uses the mean operation instead of the attention mechanism in Eqs. (6), (7), (10) and (11), where the enhanced modeling methods are changed while the enhanced data remains unchanged.
- **Data Variants** With the enhanced modeling methods unchanged, we add enhanced data step by step to verify the usefulness of various types of enhanced data, including the following items:
  +***None*** does not use the enhanced information from the OKG, which is equivalent to the ConvE model.
  +$\boldsymbol{P^{\mathcal{E},\mathcal{R}}}$ adds the text descriptions of entities and relations contained in its own KG, which is equivalent to the above OERL+Basic model.
  +$\boldsymbol{P^{\mathcal{E},\mathcal{R}}}$+$\boldsymbol{C^{\mathcal{E}}}$ adds the textual connections $C^{\mathcal{E}}$ of entities on the basis of +$P^{\mathcal{E},\mathcal{R}}$.
  +$\boldsymbol{P^{\mathcal{E},\mathcal{R}}}$+$\boldsymbol{C^{\mathcal{R}}}$ adds the textual connections $C^{\mathcal{R}}$ of relations on the basis of +$P^{\mathcal{E},\mathcal{R}}$.
  +$\boldsymbol{P^{\mathcal{E},\mathcal{R}}}$+$\boldsymbol{C^{\mathcal{E}}}$+$\boldsymbol{C^{\mathcal{R}}}$ adds the textual connections $C^{\mathcal{E}}$, $C^{\mathcal{R}}$ of both entities and relations on the basis of +$P^{\mathcal{E},\mathcal{R}}$.
  +$\boldsymbol{P^{\mathcal{E},\mathcal{R}}}$+$\boldsymbol{C^{\mathcal{E}}}$+$\boldsymbol{S^{\mathcal{E}}}$ adds the structural connection $S^{\mathcal{E}}$ on the basis of +$P^{\mathcal{E},\mathcal{R}}$+$C^{\mathcal{E}}$.
  +$\boldsymbol{P^{\mathcal{E},\mathcal{R}}}$+$\boldsymbol{C^{\mathcal{E}}}$+$\boldsymbol{C^{\mathcal{R}}}$+$\boldsymbol{S^{\mathcal{E}}}$ adds the structural connection $S^{\mathcal{E}}$ on the basis of +$P^{\mathcal{E},\mathcal{R}}$+$C^{\mathcal{E}}$+$C^{\mathcal{R}}$.

In addition, we design some OERL variants of interests: textual rule extraction, synonym searching, textual semantic filtering, structural ego extraction. Please refer to the ablation study in §5.5 for detailed settings.

5.3 Settings

The general baseline models, used to capture structural features of KGs, are reproduced with open-source implementations [5] [6]. The variants and standard OERL model are implemented with our code. where the optimizer is set to Adam, entity and relation representations are initialized randomly, Enc module is set ot BiGRU, word vectors are initialized with the pretrained GloVe embeddings, the representation dimension is set to 300, initial learning rate is set to 5e-5, dropout is set to 0.5, LeakyRelu alpha is set to 0.2, and the maximum number of entity-mentions, relation-mentions, and triples used to enhance the KG is set to 32.

We use classical evaluation metrics to evaluate the models: the proportion of correct test triples ranked in the top N predictions (H@N), mean reciprocal rank (MRR) and mean rank (MR). All models are evaluated with the filter setting. A model with better performance should have lower MR $\downarrow$, higher MRR $\uparrow$ and higher H@N $\uparrow$.

5.4 Results

In this section, experiments are presented to prove the superiority of the proposed OERL: (1) Performance w.r.t.

5. https://github.com/thunlp/OpenKE
6. https://github.com/feiwangyuzhou/GGAE

This article has been accepted for publication in IEEE Transactions on Knowledge and Data Engineering. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TKDE.2022.3218850

TABLE 2
Experimental results of General Baselines, Model Variants and the proposed OERL model on the *CN100K* dataset, where Head, Tail and Avg columns show the results of predicting the head entity, tail entity and the average of both.

Model	MR ↓			MRR ↑			H@1 ↑			H@10 ↑			H@100 ↑		
	Head	Tail	Avg	Head	Tail	Avg	Head	Tail	Avg	Head	Tail	Avg	Head	Tail	Avg
TransE	3906.0	6353.1	5129.6	16.27	17.11	16.69	2.33	2.00	2.17	41.67	43.92	42.79	67.83	67.50	67.67
TransH	3625.5	6202.8	4914.2	15.35	16.17	15.76	1.33	1.00	1.17	40.58	44.33	42.46	66.33	65.25	65.79
TransD	3820.3	6407.3	5113.8	14.97	15.02	14.99	1.67	1.67	1.67	37.83	39.67	38.75	63.17	64.00	63.58
DistMult	3655.4	6822.6	5239.0	11.86	9.83	10.84	4.83	3.83	4.33	25.17	22.17	23.67	59.17	55.50	57.33
ComlEx	5242.9	7258.4	6250.7	14.83	13.16	13.99	6.83	5.58	6.21	31.67	28.42	30.04	60.25	57.42	58.83
SimplE	3855.8	7028.4	5442.1	11.87	10.53	11.20	4.58	4.25	4.42	27.08	22.00	24.54	60.25	54.67	57.46
RotatE	2729.2	6820.5	4774.9	20.06	21.99	21.02	6.92	8.50	7.71	44.33	48.42	46.38	71.67	73.17	72.42
ConvE	2246.2	**3040.4**	2643.3	21.04	23.78	22.41	13.00	14.33	13.67	37.50	42.42	39.96	66.17	66.00	66.08
KBGAT	3717.8	3428.9	3573.4	19.32	18.78	19.05	12.08	11.50	11.79	52.83	51.83	52.33	61.17	61.08	61.13
GGAE	2923.1	3729.2	3326.2	20.29	21.06	20.67	13.42	14.00	13.71	51.92	53.92	52.92	61.58	62.08	61.83
OERL+Basic	**735.5**	3232.0	**1983.7**	30.58	32.04	31.31	19.00	21.67	20.33	54.58	52.42	53.50	85.25	72.42	78.83
OERL+MLP	1229.8	5031.7	3130.7	31.25	33.16	32.20	20.75	22.33	21.54	54.08	55.58	54.83	83.00	75.50	79.25
OERL+Mean	1176.1	5100.4	3138.2	34.75	39.25	37.00	22.33	28.50	25.42	60.50	60.67	60.58	86.75	77.50	82.13
OERL(Our)	1082.2	4698.7	2890.5	**36.57**	**40.33**	**38.45**	**24.42**	**29.08**	**26.75**	**63.00**	**62.25**	**62.63**	**88.50**	**77.67**	**83.08**

TABLE 3
Experimental results of General Baselines, Model Variants and the proposed OERL model on the *OEKG* dataset, where Head, Tail and Avg columns show the results of predicting the head entity, tail entity and the average of both.

Model	MR ↓			MRR ↑			H@1 ↑			H@10 ↑			H@100 ↑		
	Head	Tail	Avg	Head	Tail	Avg	Head	Tail	Avg	Head	Tail	Avg	Head	Tail	Avg
TransE	11772.6	10983.7	11378.1	2.82	3.54	3.18	1.78	1.78	1.78	4.60	6.59	5.59	11.95	19.55	15.75
TransH	12005.8	10413.4	11209.6	2.74	3.24	2.99	1.78	1.78	1.78	4.49	5.71	5.10	10.94	17.39	14.16
TransD	12881.6	11917.1	12399.3	2.69	3.06	2.88	1.78	1.78	1.78	4.43	5.44	4.93	10.24	15.61	12.93
DistMult	11688.1	9231.6	10459.9	0.93	1.38	1.16	0.35	0.42	0.38	1.67	2.86	2.26	7.32	12.26	9.79
ComlEx	13764.3	11629.9	12697.1	1.10	1.32	1.21	0.59	0.63	0.61	1.88	2.23	2.06	5.96	9.30	7.63
SimplE	12333.2	9764.0	11048.6	0.86	1.22	1.04	0.35	0.49	0.42	1.50	2.33	1.92	5.92	10.56	8.24
RotatE	**6896.2**	3936.8	5416.5	3.13	3.71	3.42	1.74	2.13	1.93	5.33	5.75	5.54	15.12	21.46	18.29
ConvE	8311.3	4344.8	6328.0	2.35	6.01	4.18	1.25	3.24	2.25	4.36	10.70	7.53	12.40	29.02	20.71
KBGAT	14869.7	13171.2	14020.4	1.47	1.84	1.66	0.70	0.77	0.73	6.24	8.68	7.46	8.19	11.64	9.91
GGAE	15102.5	14082.2	14592.4	2.28	2.56	2.42	1.57	1.50	1.53	6.38	8.57	7.47	8.61	12.09	10.35
OERL+Basic	7575.3	3647.4	5611.4	3.19	8.36	5.78	1.71	4.91	3.31	6.10	14.43	10.26	15.23	36.13	25.68
OERL+MLP	7043.4	3366.2	5204.8	3.64	8.05	5.85	2.06	4.36	3.21	6.48	15.47	10.98	16.90	37.11	27.00
OERL+Mean	6910.3	**3310.6**	**5110.5**	4.41	10.24	7.33	2.37	6.03	4.20	7.70	**18.54**	13.12	18.29	**41.18**	29.74
OERL(Our)	6960.6	3343.1	5151.9	**4.58**	**10.40**	**7.49**	**2.54**	**6.45**	**4.49**	**8.05**	18.26	**13.15**	**18.92**	40.98	**29.95**

Baselines: The performance improvements of OERL over general baselines and model variants. (2) Performance w.r.t. Data Variants: The performance improvements of OERL over the types of enhanced data.

- **Performance w.r.t. Baselines**

The experimental results are shown in Table 2 for *CN100K* and Table 3 for *OEKG*. The purpose of this part is to verify the superiority of the proposed OERL over general baselines and model variants from two perspectives: Q1: Is enhanced information from the OKG useful? Q2: Is the enhanced modeling method effective?

Q1: Is enhanced information from the OKG useful? Comparing the results of OERL (with enhanced information from the OKG) over general baselines (without enhanced information from the OKG) in Table 2 and Table 3, the performance of the proposed OERL exhibits a great improvement over that of state-of-the-art general baselines across the metrics (Avg score) – the MRR metric increases by 16.0 points and the H@1,10,100 metrics increase by 13.0, 9.7, 10.7 points on the *CN100K* dataset, and the MRR metric increases by 3.3 points and the H@1,10,100 metrics increase by 2.2, 5.6, 9.2 points on the *OEKG* dataset. In addition to the Avg results, the performance of predicting the head (Head) and tail (Tail) has also been greatly improved, e.g. the H@1 metric increases by 11.0 (Head), 14.7 (Tail) points on the *CN100K* dataset, increases by 0.8 (Head), 3.2 (Tail) points on the *OEKG* dataset. In another set of comparisons: OERL+Basic (with text descriptions from the KG, without enhanced information from the OKG) and OERL (with both), the performance of OERL is significantly better than that of OERL+Basic by a substantial margin across the metrics (Avg) – the MRR metric increases by 7.1 points and the H@1,10,100 metrics increase by 6.4, 9.1, 4.3 points on *CN100K*, and the MRR metric increases by 1.7 points and the H@1,10,100 metrics increase by 1.2, 2.9, 4.3 points on *OEKG*. From the above experiments and analysis, enhanced information from the OKG has a strong enhanced ability for representation learning of KGs.

Q2: Is the enhanced modeling method effective? We compare OERL+MLP, OERL+Mean and OERL, which use different enhanced modeling methods to absorb favorable enhanced information. From the results in Table 2 and

TABLE 4
Experimental results of Data Variants, where the score is the Avg version.

Enhanced Data Type	*CN100K*					*OEKG*				
	MR ↓	MRR ↑	H@1 ↑	H@10 ↑	H@100 ↑	MR ↓	MRR ↑	H@1 ↑	H@10 ↑	H@100 ↑
+*None*	2643.3	22.41	13.67	39.96	66.08	6328.0	4.18	2.25	7.53	20.71
$+P^{\mathcal{E},\mathcal{R}}$ (Basic)	1983.7	31.31	20.33	53.50	78.83	5611.4	5.78	3.31	10.26	25.68
$+P^{\mathcal{E},\mathcal{R}} + C^{\mathcal{E}}$	3021.3	36.08	24.21	60.29	82.38	5140.0	6.89	4.06	12.09	28.73
$+P^{\mathcal{E},\mathcal{R}} + C^{\mathcal{R}}$	3272.4	29.75	19.29	50.83	75.38	5276.9	6.38	3.82	11.38	28.28
$+P^{\mathcal{E},\mathcal{R}} + C^{\mathcal{E}} + C^{\mathcal{R}}$	3088.0	35.89	24.29	59.88	82.04	5105.6	7.28	4.32	13.05	29.81
$+P^{\mathcal{E},\mathcal{R}} + C^{\mathcal{E}} + S^{\mathcal{E}}$	2890.5	38.45	26.75	62.63	83.08	5166.0	6.84	3.97	12.35	29.55
$+P^{\mathcal{E},\mathcal{R}} + C^{\mathcal{E}} + C^{\mathcal{R}} + S^{\mathcal{E}}$	2724.6	36.61	24.96	60.29	81.25	5079.7	7.36	4.49	13.10	29.67

Table 3, OERL with the attention method outperforms OERL+MLP and OERL+Mean in all metrics on the *CN100K* dataset and most metrics on the *OEKG* dataset. Taking the *CN100K* dataset as an example, the H@1 metric of OERL increases by 3.7 (Head), 6.8 (Tail), 5.2 (Avg) compared to that of OERL+MLP, and increases by 2.1 (Head), 0.6 (Tail), 1.3 (Avg) compared to that of OERL+Mean. This shows that the proposed OERL model has the ability to select favorable information and eliminate noise data.

Overall, through innovative OKG-enhanced information and effective modeling methods, OERL achieves substantial improvements in comparison to all baselines. These powerful experiments and analyses prove the superiority of the OERL model and the feasibility of the OKG enhanced-strategy. Next, we present detailed experiments of the enhanced data type to further confirm this effectiveness.

- **Performance w.r.t. Data Variants**

The experimental results of data variants with different types of enhanced data are shown in Table 4. The purpose of this part is to introduce the performance improvements of OERL over the types of enhanced data. With unchanged enhanced modeling methods, we add enhanced data step by step to verify the usefulness of various types of enhanced data. We conduct detailed analysis from three angles: Q1: Is text descriptions in its own KG useful? Q2: Are textual connections from the OKG useful? Q3: Is structural connection from the OKG useful?

Q1: Is text descriptions in its own KG useful? Comparing the results of +*None* without text descriptions and $+P^{\mathcal{E},\mathcal{R}}$ with text descriptions, the performance of $+P^{\mathcal{E},\mathcal{R}}$ shows a great improvement over that of +*None*. In particular, compared to +*None*, the MRR and H@1,10,100 metrics of $+P^{\mathcal{E},\mathcal{R}}$ increases by 8.9, 6.7, 13.5, 12.8 points on *CN100K*, and 1.6, 1.1, 2.7, 5.0 points on *OEKG*. These significant effects prove the usefulness of text descriptions in its own KG.

Q2: Are textual connections from the OKG useful? Textual connections include $C^{\mathcal{E}}$ of entities and $C^{\mathcal{R}}$ of relations, so we present the analysis for just $C^{\mathcal{E}}$, just $C^{\mathcal{R}}$ and both of them. For just $C^{\mathcal{E}}$, we compare the results of $+P^{\mathcal{E},\mathcal{R}}$ (without $C^{\mathcal{E}}$) and $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}$ (with $C^{\mathcal{E}}$). $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{R}}$ outperforms $+P^{\mathcal{E},\mathcal{R}}$ with strong improvements on all the metrics, notably 4.8 points in the ARR metric and 3.9, 6.8, 3.5 points in the H@1,10,100 metrics on the *CN100K* dataset, and 1.1 point in the ARR metric and 0.7, 1.8, 3.1 points in the H@1,10,100 metrics on the *OEKG* dataset. For just $C^{\mathcal{R}}$, we compare the results of $+P^{\mathcal{E},\mathcal{R}}$ (without $C^{\mathcal{R}}$) and $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{R}}$ (with $C^{\mathcal{R}}$). $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{R}}$ outperforms $+P^{\mathcal{E},\mathcal{R}}$ by a substantial margin across the metrics – the ARR metric increases by 0.6 points and the H@1,10,100 metrics increase by 0.5, 1.1, 2.6 points on the *OEKG* dataset, but do not show improvements on the *CN100K* dataset. Through the observation of the dataset, *CN100K* has only 34 relations, and the relation type is universal, e.g., *is*, *has*. These fewer relations and universal features are the main reason for the low expressiveness of $C^{\mathcal{R}}$ with respect to the *CN100K* dataset. For both $C^{\mathcal{E}}$ and $C^{\mathcal{R}}$, the performance of $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}+C^{\mathcal{R}}$ (with both $C^{\mathcal{E}}$ and $C^{\mathcal{R}}$) is better than that of $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}$ (with just $C^{\mathcal{E}}$) and $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{R}}$ (with just $C^{\mathcal{R}}$) with respect to the *OEKG* dataset. Thus far, the above experimental results and analysis prove the strong enhancing ability of textual connections $C^{\mathcal{E}}$, $C^{\mathcal{R}}$ of entities and relations.

Q3: Is structural connection from the OKG useful? $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}+S^{\mathcal{E}}$ and $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}+C^{\mathcal{R}}+S^{\mathcal{E}}$ are two models with structural connections $S^{\mathcal{E}}$, which are compared to $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}$ and $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}+C^{\mathcal{R}}$, respectively. On the *CN100K* dataset, because of the above futility of $C^{\mathcal{R}}$, we compare $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}$ (without $S^{\mathcal{E}}$) and $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}+S^{\mathcal{E}}$ (with $S^{\mathcal{E}}$). $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}+S^{\mathcal{E}}$ enhanced with structural connections $S^{\mathcal{E}}$, achieves great improvements on all metrics versus $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}$. Among them, the ARR metric increases by 2.4 points and the H@1,10,100 metrics increase by 2.5, 2.3, 0.7 points. On the *OEKG* dataset, comparing ($+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}+C^{\mathcal{R}}$, $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}+C^{\mathcal{R}}+S^{\mathcal{E}}$), or ($+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}$, $+P^{\mathcal{E},\mathcal{R}}+C^{\mathcal{E}}+S^{\mathcal{E}}$), the performance of models enhanced with $S^{\mathcal{E}}$ is better than that without $S^{\mathcal{E}}$. Therefore, structural connections from the OKG are effective in enhancing representation learning of the KG.

Through the experimental results and analysis of the enhanced data type, both textual connections and structural connection from OKGs are able to improve the performance of representation learning of KGs.

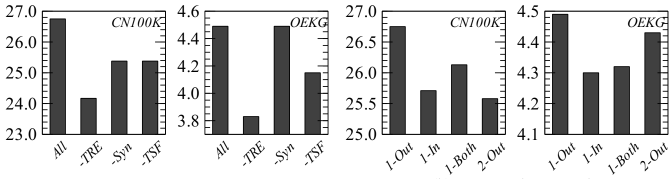
5.5 Ablation Study

To fully verify the effectiveness of each component of OERL, we present the ablation study shown in Fig. 5.

- **Components of Textual Connection**

This part presents the ablation study of textual rule extraction (TRE), synonym searching (Syn) and textual semantic filtering (TSF) in the Textual Connection section (§4.1), where textual connections and structural connections are changed, enhanced modeling methods remain unchanged. The compared models in this part include the following:

All uses textual rule extraction with synonym searching to find related mentions, and then uses textual semantic filtering to filter noise.



(a) Textual Connection (b) Structural Connection

Fig. 5. Results of ablation study over textual and structural Connections.

-TRE randomly extracts related mentions from the OKG instead of *All*. Subsequently, the structural connections are also randomly selected.

-Syn removes the synonym searching module from *All*.

-TSF removes the textual semantic filtering module, and the structural connection is extracted based on the no-filtering textual connections.

Fig. 5(a) shows the results of the above models. Comparing the results of *All* and *-TRE*, the performance of *All* is better than that of *-TRE* on both datasets, which proves the effectiveness of the proposed searching method over random ones. Comparing the results of *All* and *-Syn*, we observe that *All* outperforms *-Syn* on *CN100K*, but exhibits equal performance with *-Syn* on the *OEKG*. The motivation for the synonym searching module is to expand the search scope and find as many related mentions as possible. After statistical analysis of the *OEKG* dataset, without using synonyms, the number of related mentions in $C_h^{\mathcal{E}}$, $C_r^{\mathcal{R}}$ has been large enough, which makes synonyms unable to play a role. Moreover, this semantic filtering module is important and indispensable in filtering noise, because *-TSF* without the semantic filtering module shows poor performance than *All* which uses the semantic filtering. Overall, the proposed textual rule extraction, synonym searching and textual semantic filtering can help entities and relations obtain their related entity-mentions and relation-mentions from OKGs.

- **Components of Structural Connection**

This part presents the ablation study of structural ego extraction in Structural Connection section (§4.2), where structural connections are changed, while textual connections and enhanced modeling methods remain unchanged. The performance improvements of the proposed OERL over the different scope of structures are evaluated:

SEE-1-Out uses outgoing triples of the 1-ego subgraph.

SEE-1-In uses incoming triples of the 1-ego subgraph.

SEE-1-Both uses both outgoing and incoming triples of the 1-ego subgraph.

SEE-2-Out uses outgoing triples of the 2-ego subgraph.

Fig. 5(b) shows the results of structural ego extraction. Comparing the three models (*SEE-1-Out*, *SEE-1-In*, *SEE-1-Both*), we observe that *SEE-1-Out* with just outgoing triples achieves better performance than the others. Therefore, outgoing triples are more effective in improving the performance of representation learning. We also attempt to expand the scope from the 1-ego subgraph to the 2-ego subgraph. The performance of *SEE-2-Out* with the 2-ego subgraph does not improve compared to that of *SEE-1-Out*. After analysis, the scale and noise of 2-ego increase significantly compared with those of 1-ego subgraph, which also brings great computational pressure to the model.

TABLE 5
Two triples to be predicted, where Triple1 is from CN100K and Triple2 is from OEKG. TC, SC denote textual and structural connections. Rank denotes the position of the correct answer, the smaller the better.

Triple1	(*fish*, *at location*, ?) Correct answer: *water*
TC	$C^{\mathcal{E}}$: *fishes*, *all fish*, *tropical fish*. $C^{\mathcal{R}}$: *at were located*, *as of is located*, *was used as a location for*.
SC	[*pools*, *be populated with*, *fish*] [*fish*, *is*, *meat*] [*eel*, *is*, *fish*]
Rank	*All*: 3 *-SC*: 6 *-SCTC*: 14
Triple2	(*legislature*, *convene*, ?) Correct answer: *session*
TC	$C^{\mathcal{E}}$: *a legislature*, *provincial legislature*, *alberta legislature*. $C^{\mathcal{R}}$: *convene on*, *convene in*, *convenes*.
SC	[*legislature*, *has*, *first act*] [*legislature*, *has*, *approval*] [*legislature*, *permit*, *divorce*]
Rank	*All*: 21 *-SC*: 29 *-SCTC*: 106

Through the above ablation study, each component of OERL is important and indispensable in enhancing the representation learning of KGs.

- **Ablation Study of Cases**

This part presents some cases to show the advantages of fusing textual and structural connections. The compared models in this part include the following:

All is our full model.

-SC removes the structural connection.

-SCTC removes both structural and textual connections.

Table 5 shows two triples which are predicted by *All*, *-SC*, *-SCTC*, respectively. The rank scores of *All*, *-SC*, *-SCTC* are 3, 4, 14 for triple1 and 21, 29, 106 for triple2, which indicate that *All* enhanced with both textual and structural connections performs well than *-SC* and *-SCTC*. To be specific, a model is required to predict the tail entity *water* according to head entity *fish* and relation *at location* in triple1. The textual connections *fishes*, *all fish*, *tropical fish* in $C^{\mathcal{E}}$ of head entity *fish* can supply various sufficient descriptions to help the understanding of the entity, and similarly for the relation *at location*. The structural connection [*pools*, *be populated with*, *fish*] provides favorable evidence for the selection of the correct answer *water* because the pools contain water. Through the above qualitative analysis of cases, textual and structural connections are capable to provide supplementary information to help the representation learning of KGs.

6 CONCLUSION

In this paper, we propose a new OERL, open knowledge graph enhanced representation learning of knowledge graphs. OERL extracts textual and structural connections between KGs and OKGs, models and transfers the enhanced features from the connections to representation learning of KGs. Experimental verification and analysis prove the feasibility and superiority of the proposed OERL. Moreover, the proposed open knowledge graph enhanced strategy is not limited to the link prediction task, it can also be used in other downstream tasks, e.g., KG-QA and KG-Dialog. Overall, we expect more researchers to devote attention to the proposed open knowledge graph enhanced strategy, and hope that the proposed strategy and model can help representation learning reach a higher level.

ACKNOWLEDGMENT

The project is supported by National Natural Science Foundation of China (62172086, 61872074, 62106039), Fundamental Research Funds for the Central Universities of China (N2016008), and the Chinese Scholarship Council.

REFERENCES

[1] A. Bordes, N. Usunier, A. García-Durán, J. Weston, and O. Yakhnenko, "Translating embeddings for modeling multi-relational data," in *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems, December 5-8, Lake Tahoe, Nevada, United States*, 2013, pp. 2787–2795.

[2] T. Dettmers, P. Minervini, P. Stenetorp, and S. Riedel, "Convolutional 2d knowledge graph embeddings," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI, New Orleans, Louisiana, USA, February 2-7*, 2018, pp. 1811–1818.

[3] Q. Li, D. Wang, S. Feng, C. Niu, and Y. Zhang, "Global graph attention embedding network for relation prediction in knowledge graphs," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2021.

[4] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, "Heterogeneous graph attention network," in *The World Wide Web Conference, WWW, San Francisco, CA, USA, May 13-17*, 2019, pp. 2022–2032.

[5] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph embedding by translating on hyperplanes," in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, Québec City, Québec, Canada*, 2014, pp. 1112–1119.

[6] G. Ji, S. He, L. Xu, K. Liu, and J. Zhao, "Knowledge graph embedding via dynamic mapping matrix," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics, ACL, July 26-31, Beijing, China, Volume 1: Long Papers*, 2015, pp. 687–696.

[7] S. M. Kazemi and D. Poole, "Simple embedding for link prediction in knowledge graphs," in *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems, NeurIPS, December 3-8, Montréal, Canada*, 2018, pp. 4289–4300.

[8] B. Yang, W. Yih, X. He, J. Gao, and L. Deng, "Embedding entities and relations for learning and inference in knowledge bases," in *3rd International Conference on Learning Representations, ICLR, San Diego, CA, USA, May 7-9, Conference Track Proceedings*, 2015.

[9] T. Trouillon, J. Welbl, S. Riedel, É. Gaussier, and G. Bouchard, "Complex embeddings for simple link prediction," in *Proceedings of the 33nd International Conference on Machine Learning, ICML, New York City, NY, USA, June 19-24*, vol. 48, 2016, pp. 2071–2080.

[10] Z. Sun, Z. Deng, J. Nie, and J. Tang, "Rotate: Knowledge graph embedding by relational rotation in complex space," in *7th International Conference on Learning Representations, ICLR,New Orleans, LA, USA, May 6-9*, 2019.

[11] S. Guo, Q. Wang, B. Wang, L. Wang, and L. Guo, "SSE: semantically smooth embedding for knowledge graphs," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 4, pp. 884–897, 2017.

[12] R. Xie, Z. Liu, and M. Sun, "Representation learning of knowledge graphs with hierarchical types," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI, New York, NY, USA, 9-15 July*, 2016, pp. 2965–2971.

[13] R. Xie, Z. Liu, J. Jia, H. Luan, and M. Sun, "Representation learning of knowledge graphs with entity descriptions," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, Phoenix, Arizona, USA*, 2016, pp. 2659–2665.

[14] Z. Wang and J. Li, "Text-enhanced representation learning for knowledge graph," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI, New York, NY, USA, 9-15 July*, 2016, pp. 1293–1299.

[15] B. An, B. Chen, X. Han, and L. Sun, "Accurate text-enhanced knowledge graph representation learning," in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, New Orleans, Louisiana, USA, June 1-6, Volume 1 (Long Papers)*, 2018, pp. 745–755.

[16] L. Yao, C. Mao, and Y. Luo, "KG-BERT: BERT for knowledge graph completion," *CoRR*, vol. abs/1909.03193, 2019.

[17] Z. Zhang, X. Liu, Y. Zhang, Q. Su, X. Sun, and B. He, "Pretrain-kge: Learning knowledge representation from pretrained language models," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings, EMNLP, Online Event, 16-20 November*, 2020, pp. 259–266.

[18] S. Broscheit, K. Gashteovski, Y. Wang, and R. Gemulla, "Can we predict new facts with open knowledge graph embeddings? A benchmark for open link prediction," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL, Online, July 5-10*, 2020, pp. 2296–2308.

[19] K. Gashteovski, S. Wanner, S. Hertling, S. Broscheit, and R. Gemulla, "OPIEC: an open information extraction corpus," in *1st Conference on Automated Knowledge Base Construction, AKBC, Amherst, MA, USA, May 20-22*, 2019.

[20] S. Lv, D. Guo, J. Xu, D. Tang, N. Duan, M. Gong, L. Shou, D. Jiang, G. Cao, and S. Hu, "Graph-based reasoning over heterogeneous external knowledge for commonsense question answering," in *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI, New York, NY, USA, February 7-12*, 2020, pp. 8449–8456.

[21] H. Zhang, Z. Liu, C. Xiong, and Z. Liu, "Grounded conversation generation as guided traverses in commonsense knowledge graphs," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL, Online, July 5-10*, 2020, pp. 2031–2043.

[22] Y. Xu, Y. Zhu, Y. Shen, and J. Yu, "Learning shared vertex representation in heterogeneous graphs with convolutional networks for recommendation," in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI, Macao, China, August 10-16*, 2019, pp. 4620–4626.

[23] Y. Lin, Z. Liu, M. Sun, Y. Liu, and X. Zhu, "Learning entity and relation embeddings for knowledge graph completion," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, Austin, Texas, USA*, 2015, pp. 2181–2187.

[24] M. Nickel, V. Tresp, and H. Kriegel, "Factorizing YAGO: scalable machine learning for linked data," in *Proceedings of the 21st World Wide Web Conference, WWW, Lyon, France, April 16-20*, 2012, pp. 271–280.

[25] R. Socher, D. Chen, C. D. Manning, and A. Y. Ng, "Reasoning with neural tensor networks for knowledge base completion," in *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems, December 5-8, Lake Tahoe, Nevada, United States*, 2013, pp. 926–934.

[26] Z. Wang, J. Zhang, J. Feng, and Z. Chen, "Knowledge graph and text jointly embedding," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, October 25-29, Doha, Qatar*, 2014, pp. 1591–1601.

[27] T. Long, R. Lowe, J. C. K. Cheung, and D. Precup, "Leveraging lexical resources for learning entity embeddings in multi-relational data," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL, August 7-12, Berlin, Germany, Volume 2: Short Papers*, 2016.

[28] O. Etzioni, A. Fader, J. Christensen, S. Soderland, and Mausam, "Open information extraction: The second generation," in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence, IJCAI, Barcelona, Catalonia, Spain, July 16-22*, 2011, pp. 3–10.

[29] A. Fader, S. Soderland, and O. Etzioni, "Identifying relations for open information extraction," in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, EMNLP, 27-31 July, John McIntyre Conference Centre, Edinburgh, UK*, 2011, pp. 1535–1545.

[30] Mausam, M. Schmitz, S. Soderland, R. Bart, and O. Etzioni, "Open language learning for information extraction," in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing, EMNLP, July 12-14, Jeju Island, Korea*, 2012, pp. 523–534.

[31] L. D. Corro and R. Gemulla, "Clausie: clause-based open information extraction," in *22nd International World Wide Web Conference, WWW, Rio de Janeiro, Brazil, May 13-17*, 2013, pp. 355–366.

[32] H. Pal and Mausam, "Demonyms and compound relational nouns in nominal open IE," in *Proceedings of the 5th Workshop on Automated Knowledge Base Construction, AKBC@NAACL-HLT, San Diego, CA, USA, June 17*, 2016, pp. 35–39.
[33] S. Saha, H. Pal, and Mausam, "Bootstrapping for numerical open IE," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL, Vancouver, Canada, July 30 - August 4, Volume 2: Short Papers*, 2017, pp. 317–323.
[34] S. Saha and Mausam, "Open information extraction from conjunctive sentences," in *Proceedings of the 27th International Conference on Computational Linguistics, COLING, Santa Fe, New Mexico, USA, August 20-26*, 2018, pp. 2288–2299.
[35] L. A. Galárraga, C. Teflioudi, K. Hose, and F. M. Suchanek, "AMIE: association rule mining under incomplete evidence in ontological knowledge bases," in *22nd International World Wide Web Conference, WWW, Rio de Janeiro, Brazil, May 13-17*, 2013, pp. 413–422.
[36] L. Galárraga, G. Heitz, K. Murphy, and F. M. Suchanek, "Canonicalizing open knowledge bases," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, CIKM, Shanghai, China, November 3-7*, 2014, pp. 1679–1688.
[37] S. Vashishth, P. Jain, and P. P. Talukdar, "CESI: canonicalizing open knowledge bases using embeddings and side information," in *Proceedings of the 2018 World Wide Web Conference, WWW, Lyon, France, April 23-27*, 2018, pp. 1317–1327.
[38] T. Jiang, T. Zhao, B. Qin, T. Liu, N. V. Chawla, and M. Jiang, "Canonicalizing open knowledge bases with multi-layered metagraph neural network," *CoRR*, vol. abs/2006.09610, 2020.
[39] S. Gupta, S. Kenkre, and P. P. Talukdar, "Care: Open knowledge graph embeddings," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, EMNLP, Hong Kong, China, November 3-7*, 2019, pp. 378–388.
[40] G. A. Miller, *WordNet: An electronic lexical database*. MIT press, 1998.
[41] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, 2019, pp. 4171–4186.
[42] K. Cho, B. van Merrienboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP, October 25-29, Doha, Qatar*, 2014, pp. 1724–1734.
[43] R. Speer and C. Havasi, "Conceptnet 5: A large semantic network for relational knowledge," in *The People's Web Meets NLP*. Springer, 2013, pp. 161–176.
[44] D. Nathani, J. Chauhan, C. Sharma, and M. Kaul, "Learning attention-based embeddings for relation prediction in knowledge graphs," in *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL, Florence, Italy, July 28- August 2, Volume 1: Long Papers*, 2019, pp. 4710–4723.
[45] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in *6th International Conference on Learning Representations, ICLR, Vancouver, BC, Canada, April 30 - May 3, Conference Track Proceedings*, 2018.

Qian Li received an M.S. degree from the School of Computer Science and Engineering, Northeastern University, Shenyang, China, in June 2018.

She has been a Ph.D. candidate at the School of Computer Science and Engineering, Northeastern University, Shenyang, China, since September 2018. She has authored or coauthored many articles in the top and major tiered journals and conferences, including International Joint Conferences on Artificial Intelligence (IJCAI), Annual Meeting of the Association for Computational Linguistics (ACL), Conference on Empirical Methods in Natural Language Processing (EMNLP), and IEEE Transactions on Neural Networks and Learning Systems (TNNLS). Her research interests include knowledge graph representation learning and natural language processing.

Daling Wang received a Ph.D. degree in computer software and theory from Northeastern University, Shenyang, China, in March 2003.

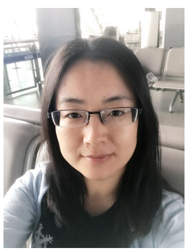
She is a Professor with the School of Computer Science and Engineering, Northeastern University, Shenyang, China. She has authored or coauthored many articles in the top and major tiered journals and conferences, including International Joint Conferences on Artificial Intelligence (IJCAI), Annual Meeting of the Association for Computational Linguistics (ACL), AAAI Conference on Artificial Intelligence (AAAI), International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), The Web Conference (WWW), Conference on Empirical Methods in Natural Language Processing (EMNLP), and IEEE Transactions on Neural Networks and Learning Systems (TNNLS). Her research interests include social media processing, sentiment analysis, data mining, and information retrieval.

Shi Feng received a Ph.D. degree in computer software and theory from Northeastern University, Shenyang, China, in January 2011.

He is an Associate Professor with the School of Computer Science and Engineering, Northeastern University, Shenyang, China. He has authored or coauthored more than 20 articles in top-tier journals and conferences, including International Joint Conferences on Artificial Intelligence (IJCAI), Annual Meeting of the Association for Computational Linguistics (ACL), AAAI Conference on Artificial Intelligence (AAAI), International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), The Web Conference (WWW), and Conference on Empirical Methods in Natural Language Processing (EMNLP). His research interests include sentiment analysis and dialogue systems.

Kaisong Song received a Ph.D. degree in computer software and theory from Northeastern University, Shenyang, China, in July 2017.

He is an algorithm expert in DAMO academy, Alibaba Group, Hangzhou, China. He has authored or coauthored more than 20 articles in top-tier journals and conferences, including International Joint Conferences on Artificial Intelligence (IJCAI), AAAI Conference on Artificial Intelligence (AAAI), International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR), The Web Conference (WWW), and Conference on Empirical Methods in Natural Language Processing (EMNLP). His research interests include recommender systems, sentiment analysis and content security.



Yifei Zhang received a Ph.D. degree in computer software and theory from Northeastern University, Shenyang, China, in July 2009.

She is an Assistant Professor with the School of Computer Science and Engineering, Northeastern University, Shenyang, China. Her research interests include image processing and machine learning.

Ge Yu received a Ph.D degree in Computer Science from Kyushu University of Japan, in 1996.

He has been a full professor at Northeastern University of China since 1996. He serves as the associate editor of Chinese Journal of Computers, Journal of Software, and Journal of Computer Research and Development. He is a CCF fellow, and an IEEE senior member and an ACM member. His current research interests include database theory and technology, distributed and parallel systems, cloud computing and big data management, blockchain techniques and their applications.