applied sciences

MDPI

Article

An Effective Knowledgeable Label-Aware Approach for Sentential Relation Extraction

Binling Nie* **and Yiming Shao**

Digital Media School, Hangzhou Dianzi University, Hangzhou 310000, China
* Correspondence: binlingnie@hdu.edu.cn

Abstract: In recent years, sentential relation extraction has made remarkable progress with text and knowledge graphs (KGs). However, existing architectures ignore the valuable information contained in relationship labels, which KGs provide and can complement the model with additional signals and prior knowledge. To address this limitation, we propose a neural architecture that leverages knowledgeable labels to enhance sentential relation extraction. We name our proposed method knowledge label-aware sensitive relation extraction (KLA-SRE). To achieve this, we combine pre-trained static knowledge graph embeddings with learned semantic embeddings from other tokens to efficiently represent relation labels. By combining static pre-trained graph embeddings with learned word embeddings, we mitigate the inconsistency between predicted relations and given entities. Experimental results on various relation extraction benchmarks in different fields show that knowledge labels improve the F1 score by 1.6% and 1.1% on average over the baseline on standard- and minority-shot benchmarks, respectively.

Keywords: sentential relation extraction; knowledge graph; label-aware

check for updates

Citation: Nie, B.; Shao, Y. An Effective Knowledgeable Label-Aware Approach for Sentential Relation Extraction. *Appl. Sci.* **2023**, *13*, 11929. https://doi.org/10.3390/app132111929

Academic Editor: Tobias Meisen

Received: 25 September 2023
Revised: 19 October 2023
Accepted: 26 October 2023
Published: 31 October 2023

Copyright: © 2023 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

Machine learning researchers now have access to vast amounts of structured and unstructured data, including structured knowledge graphs (KGs) such as Wikidata [1], Yago [2], and UMLS [3]. KGs organize various types of information, improve the accuracy and efficiency of machine learning algorithms, and provide valuable resources for natural language processing (NLP) tasks such as entity recognition, relationship extraction, and language modeling. Advances in NLP have led to the development of sophisticated models that can perform complex tasks such as text classification and sentiment analysis on unstructured text data. As a result, the intersection of KGs and NLP is a popular research area for developing accurate, efficient, and complex models that can process structured and unstructured data.

Two particular types of work use knowledge base information to improve the classification performance of relation extraction (RE). One prominent way to leverage knowledge bases to enhance RE is through distant supervision [4]. Distantly supervised RE models evaluate entity pairs in a sentence by verifying whether they are linked in the KGs. While KGs are a valuable source of training data for RE, their usefulness goes beyond that. KGs also provide information and structures that can help resolve issues such as data sparsity and inconsistency in other tagged sources. Another increasingly popular class of methods involves integrating unstructured text data with structured information to improve the performance of natural language processing tasks, including RE [5–7]. These methods aim to improve the performance of RE tasks by providing more representative embeddings that capture the complementary information from both unstructured text and structured data. Methods such as RECON [5] dynamically select relevant information from knowledge graphs and sentences through joint representation, enabling a more complete understanding of the input. However, this knowledge is derived from entity attributes

and other information obtained from the knowledge base, rather than directly from the graph structure. Furthermore, embeddings are only trained for downstream tasks, ignoring the entire KGs. To address this, Papaluca et al. [8] incorporated KB embeddings based on precomputed topology, improving their model's performance to levels comparable to those achieved by more advanced models on typical benchmark datasets.

The existing research has not adequately explored how to provide RE models with informative and informed label representations. Incorporating topological structure information from relationship label knowledge graph embeddings can improve the overall performance of RE models in the perceptual RE setting. Relationship labels are themselves topological structures in knowledge graphs, meaning that they are semantically related to other topological structures. This correlation can be exploited to improve RE. For example, in Table 1, given the relation `father`, its corresponding triples `<zhang zun, zhang bao, father>` and `<christina stead, david george stead, father>` and related triples of tail entities `{<zhang bao, zhang zun, child>, <zhang bao, sex or gender, male>, <zhang bao, zhang, family name>}`, `{<david george stead, christina stead, child>, <david george stead, sex or gender, male>,` and `<david george stead, stead, family name>}` in its corresponding triples underly the topology structure of itself. The topology structure carries the meaning that our models can induce from knowledge graph representations. Thus, obtaining the knowledge graph embeddings of the tail entity `peter george peterson`, which are derived from its related triples `{<peter george peterson, sex or gender, male>, <peter george peterson, peterson, family name>}`, which represent the topology structure of the tail entity `peter george peterson` and are similar to that of the tail entity in train example S1, leads the model to predict the representations of `father` rather than the representations of `mother` when similar priors are used for labels and words or phrases. In addition, a relation description explains the detailed meaning of the relation and the traits of its entity pair. It helps to figure out the pivotal context of the sentence and the specific trait of the entity. For example, in Table 1, relation descriptions indicate that the subject and object of `father` indicate information about the gender of the object.

In this study, we propose a new method for knowledge label-aware sensitive relation extraction (KLA-SRE) that utilizes knowledge graph information about relational labels and combines task learning with word embeddings to improve performance. By incorporating this background knowledge, our approach outperforms existing methods in terms of accuracy and effectiveness, making it a promising avenue for future research. The KLA-SRE system consists of two modules: the first module encodes the component words and their corresponding entity graph embeddings, while the second module extracts the relations between the entities. The KLA-SRE system comprises two distinct modules, with the first module functioning as an encoder tasked with the responsibility of encoding both the constituent words and its corresponding entity graph embeddings. The second module matches the word representation derived from the first module with the information-rich relation label representation, which contains the topological structure information and the descriptive semantics of the relation label and agrees with the knowledge-enhanced sentence.

This paper presents the results of our experiments conducted on several NER datasets pertaining to diverse domains. We summarize our contribution as follows:

- We propose a model architecture that is both straightforward and efficient, drawing upon knowledgeable labels to enhance its accuracy in the field of SRE.
- Our work demonstrates the efficacy of the proposed model in low-resource environments, yielding results comparable to state-of-the-art models in high-resource contexts.

Table 1. The head entity and tail entity are indicated by red and blue, respectively. The italics in brackets are corresponding KB ids and aliases/descriptions about the relation and entity. Given the training example, it requires models to predict which relation the test example expresses. red: head entity, blue: tail entity.

Training Example:
R1: father [“id”: P22, “alias”: ”father”, “Descriptions”: “male parent of the subject”] S1: Paul Gouin the son of Lomer Gouin and granson of Honore Mercier. [“id”: “Q3371345”, “alias”: “paul gouin”, “id”: “Q764557”, “alias”: “lomer gouin”]
Test example:
S2: Holly Peterson was born in 1964 Chicago, Illinois the daughter of Peter George Peterson and psychologist Sally Peterson.
KG Topology Structure:
KG Triples of R1: ({{{“id”: “Q718095”, “alias”: “zhang zun”}, {“id”: “Q168240”, “alias”: “zhang bao”}, {“id”: P22, “alias”: “father”}, . . . , {{“id”: “Q441056”, “alias”: “christina stead”}, {“id”: “Q5234086”, “alias”: “david george stead”}, {“id”: P22, “alias”: “father”}}) Related KG Triples of zhang bao: ({{“id”: “Q168240”, “alias”: “zhang bao”}, {“id”: “Q718095”, “alias”: “zhang zun”}, {“id”: P40, “alias”: “child”}}, {{“id”: “Q168240”, “alias”: “zhang bao”}, {“id”: “Q6581097”, “alias”: “male”}{“id”: P21, “alias”: “sex or gender”}, {{“id”: “Q168240”, “alias”: “zhang bao”}, {“id”: “Q804909”, “alias”: “zhang”}, {“id”: P734, “alias”: “family name”}}) Related KG Triples of david george stead: ({{“id”: “Q5234086”, “alias”: “david george stead”}, {“id”: “Q441056”, “alias”: “christina stead”}, {“id”: P40, “alias”: “child”}}, {{“id”: “Q5234086”, “alias”: “david george stead”}, {“id”: “Q6581097”, “alias”: “male”}{“id”: P21, “alias”: “sex or gender”}, {{“id”: “Q5234086”, “alias”: “david george stead”}, {“id”: “Q55759066”, “alias”: “stead”}, {“id”: P734, “alias”: “family name”}})
KG Triples of Entities in S1: ({{“id”: “Q764557”, “alias”: “lomer gouin”}, {“id”: “Q3371345”, “alias”: “paul gouin”}, {“id”: P40, “alias”: “child”}}, {{“id”: “Q764557”, “alias”: “lomer gouin”}, {“id”: “Q6581097”, “alias”: “male”}{“id”: P21, “alias”: “sex or gender”}, {{“id”: “Q3371345”, “alias”: “paul gouin”}, {“id”: “Q16870263”, “alias”: “Gouin”}, {“id”: P734, “alias”: “family name”}}) KG Triples of Entity in S2: ({{“id”: “Q456216”, “alias”: “peter george peterson”}, {“id”: “Q6581097”, “alias”: “male”}{“id”: P21, “alias”: “sex or gender”}, {{“id”: “Q456216”, “alias”: “peter george peterson”}, {“id”: “Q12191064”, “alias”: “peterson”}, {“id”: P734, “alias”: “family name”}})

2. Related Works

2.1. Sentential Relation Extraction

Sentential relationship extraction (SRE) is a natural language processing task that aims to extract semantic relationships between entities from a single sentence. The restriction of sentential RE is that when predicting KG relationships, the document context is limited to the input sentence, excluding other occurrences of entity pairs [9,10]. Some studies continuously pretrain a pretrained language model on text with linked entities using relationship-oriented goals [10–12]. Specifically, BERT-MTB [12] proposes a matching blank objective to determine whether two relations share the same entity.

2.2. Knowledge Graph Information Enhanced RE

Recent works on knowledge graph information sentential RE mainly focus on injecting external knowledge into PLMs. Methods such as ERNIE [13] and KnowBERT [14] incorporate pretrained entity embeddings from knowledge graphs into the transformer. Some other existing models use a KG as a supplementary information source to process noisy sentences. These models, typically implemented as dual encoders (one for the text input and one for the KG input), are trained to rely more on the text encoder when given informative sentences and more on the KG encoder when faced with noisy sentences [11,15–20].

Some methods focus on dynamically learning representations of KB entities simultaneously with word representations to improve performance for sentential RE. Vashishth et al. [7]

conducted a matching procedure between the relation predicted by the Stanford OpenIE [21] pipeline and a set of relation aliases retrieved from the knowledge base. This resulted in a matched relation embedding, which was then combined with the sentence representation. Moreover, a corresponding entity-type embedding was developed using the entity type identified in the KB. The resulting entity type embedding was also concatenated to the sentence encoding. Bastos et al. [5] present a novel approach for constructing an entity attribute context embedding. The proposed method involves processing various entity properties from the knowledge base with an BiLSTM encoder [22]. In addition, triplet context embeddings are learned for each relation triplet by imposing translational properties in the embedding space [23] on the triplet and its KG neighbors. The two distinct representations are then integrated with the sentence encodings using a GP-GNN [24] and subsequently fed into a classifier for relation prediction. Nadgeri et al. [6], propose a strategy for dynamically selecting relevant information. To achieve this goal, they extract and encode numerous KB properties for each entity using a BiLSTM [22]. They then combine the resulting encodings with the sentence encoding to establish a heterogenous information graph. To obtain contextually relevant information, the graph is pruned using a combination of graph convolutional neural network techniques [25], pooling, and self-attention layers. Finally, the pruned information is aggregated and fed to the relation classifier. Andrea et al. [8] propose a novel model that builds upon existing models described by Nguyen and Verspoor [26] and Giorgi et al. [27]. The proposed model combines off-task pre-trained graph embeddings with on-task fine-tuned BERT encoders. They comprehensively evaluate the model, taking into account pre-computed knowledge base embeddings. The model is a simple extension of BERT, but it is competitive with other state-of-the-art models that directly perform on-task training on these embeddings. We namely this model PKG in our paper. However, current approaches only incorporate entity embeddings that are derived from a range of entity properties or topology-based entity embeddings, as well as information obtained from the knowledge graph. They fail to consider the topology and description of semantic information associated with relation labels. To fully utilize the benefits of knowledgeable labels in sentence-level relation extraction, this paper presents a novel model for exploiting label information within knowledge graphs.

3. Background

In this section, we establish the formulation of the relation extraction task and the few-shot RE settings.

Sentential relation extraction: A sentence, denoted by $T = (t_1, t_2, t_3, \ldots, t_l)$, is essentially a sequence of words. The entities in a sentence are represented by the set $M = m_1, m_2, \ldots, m_k$, where every $m_k = (t_i, w_2, \ldots, t_j)$ is a segment of the sentence T. These mentions are annotated using entities from a knowledge graph (KG), such that m_i:e_j, where $e_j \in E$. A relation instance is defined as $x = (T, e1, e2)$, implying the presence of a target relationship that is expected to be predicted by RE models for the given sentence's e_1 and e_2 entities. In the event that no corresponding relationship is found in the KG, the label `NA` is applied. The task of sentence-level relation extraction asserts that only the sentence containing a given pair of entities is pertinent for predicting the correct relation, while disregarding all other sentences in the set. This approach is commonly referred to as the sentential RE task.

Low-shot relation extraction: Low-shot relation extraction necessitates models that can identify previously unseen relations with minimal samples. This task involves performing relation extraction using only a limited number of annotated instances for newly introduced relations. Each few-shot relation extraction task consists of a support set, $R = r_1, \ldots, r_N$, containing annotated instances for N novel relations. Specifically, for relation i, the support set $r_i = r_i^0, \ldots, r_i^K$ contains $K, K \geq 0$ such instances. The aim is for models to acquire the ability to transfer knowledge and accurately extract the target relation, y_t, for any given instance, x, through the K-shot support set.

4. Model

In this section, we present the details of our proposed approach. Figure 1 shows the overall structure; our model consists of two modules: a knowledge-augmented sentence encoding module and a knowledge label-aware relationship classification module. The knowledge-augmented sentence-encoding module initializes sentence encodings with BERT outputs. For multi-token entities, named entity encodings are extracted and averaged. The knowledge-graph-embedding augmented model connects knowledge graph embeddings to entity encodings. Each entity is decomposed into head-to-tail representations to form candidate pairs, which are then passed to the biaffine attention layer for knowledgeable label-aware relationship classification. To ensure that the representations of entity pairs and relationship labels are in the same semantic space, the same BERT encoder is used to compute the description semantics of relationship labels. Furthermore, the topology knowledge graph embeddings of relation labels are combined into their representations for better matching abilities.

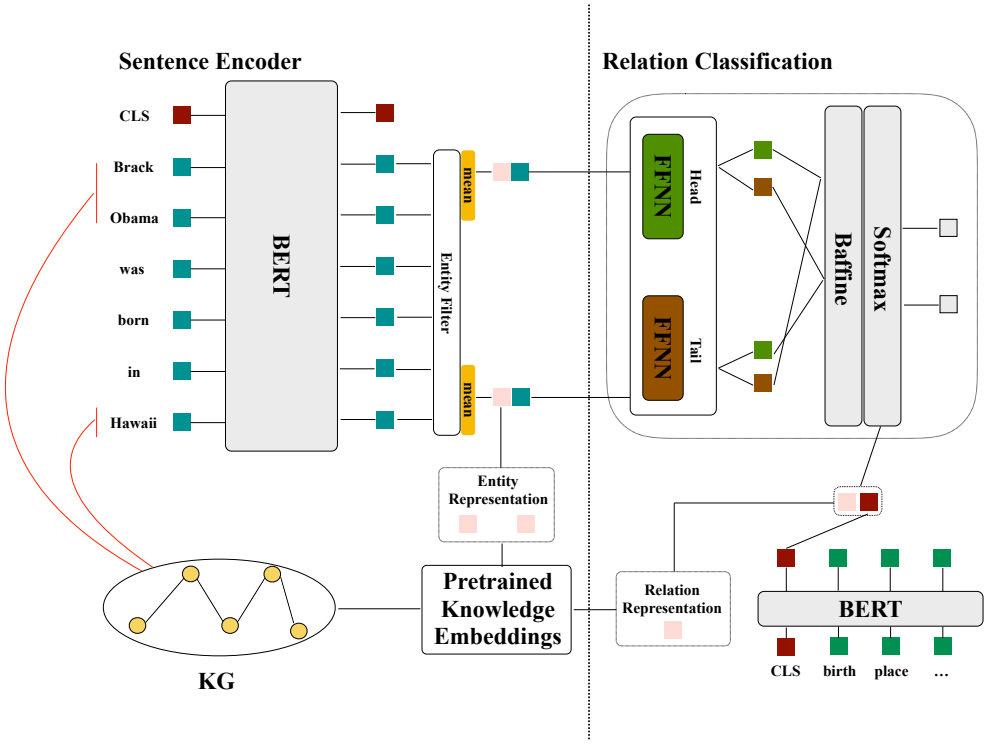


Figure 1. The proposed KLASRE, which consists of two main components: a knowledge-enhanced sentence encoder, and a knowledgeable label-aware relationship classification module. The knowledge-enhanced sentence encoder incorporates both text semantics and external knowledge to facilitate label prediction, which acts as external supplementary knowledge of sentences. The knowledgeable label-aware relation classification module utilizes the description semantics and topology structure information of relation labels to offer a knowledgeable semantic embedding, which provides supervision for label inference.

4.1. Knowledgeable Label Aware Context Encoder

Since BERT has achieved excellent performance on most NLP tasks, we adopt it as the context encoder to obtain representations of sentences and relation label descriptions.

4.1.1. Sentence Encoder

The initial encodings for the sentence's tokens are provided by the pre-trained language model. In our study, we have chosen to utilize variations of the BERT model, such as the bert-base-cased and roberta-base implementations provided by Huggingface [28]. Nevertheless, it is worth noting that any other encoder with the ability to provide context-

dependent embeddings can be utilized in place of the encoder. Throughout training, we do not restrict the encoder from modification and allow gradients to flow through it, which permits the fine-tuning of its parameters to optimize token representation for the given task. In more detail, the training procedure in our paper commences with the encoder being fully frozen and continues with unfreezing the last four layers in the initial epochs of training, as proposed in Araci's [29] work. Prior research has demonstrated that this approach does not inevitably lead to a decrease in accuracy, yet it considerably alleviates the computational load. Given a sentence, $[w_1, w_2, \ldots, w_n]$, we use BERT to generate a contextual representation of the sentence as follows:

$$X = BERT([w_1, w_2, \ldots, w_n]) \tag{1}$$

where $X \in \mathcal{R}^{n \times d_h}$.

4.1.2. Relation Description Encoder

Since our label description of the semantic encoder is identical to the sentence encoder, our architecture can accept any textual form as input to generate description representations. To ensure compatibility with prior research, we exclusively employ the natural language format of label descriptions for our primary outcomes r^{word}.

4.1.3. Collaborative Attention Module

It should be noted that in the model augmented by a graph embedding, we append the relative knowledge graph embeddings $e_i^{knowledge_{h|t}}$ and $r^{knowledge}$ of Lerer et al. [30] to represent the topology structure for head/tail entities and relation labels. As mentioned earlier, the relationship description and topology elaborate on the detailed meaning and characteristics of entity pairs. Therefore, highlighting key information in sentences is beneficial for extracting both instructional and relational features. We transform the relation representations $r^{word}, r^{knowledge}$ with W_s and W_k to obtain different representations from the perspectives of sentence and knowledge graph structures. Afterwards, to better attend to the information content of both inputs, we adopt a collaborative attention module:

$$\alpha_s = softmax(r^{word} W_s X) \tag{2}$$

$$\hat{X} = \sum_{i=1}^{n} \alpha_s X \tag{3}$$

$$\alpha_k = softmax(r^{knowledge} W_k (e^{knowledge_h} - e^{knowledge_t})) \tag{4}$$

$$e^{\hat{knowledge}_{h|t}} = \sum_{i=1}^{n} \alpha_k e^{knowledge_{h|t}} \tag{5}$$

where $\alpha_s(\alpha_k)$ is the attention weight over the word and knowledge graph representations, $X(e^{knowledge_{h|t}})$, of the sentence (topology structure) and the final weighted vector, $\hat{X}(e^{\hat{knowledge}_{h|t}})$,is obtained by the sum operation.

4.1.4. Entity Filter

The entity filter aims to eliminate non-entity tokens from text representations. For multi-token entities, two common options are to choose the last token or the average of all tokens as the identifier. We tested both options and found no significant difference in performance for our application. As a result, we chose to use the average encoding of the entity's constituent tokens, $w_i, \ldots, w_j$, as its identifier:

$$x^{word_{h|t}} = \overline{\sum_{l=i}^{j} X_l^{\hat{h}|t}} \tag{6}$$

4.2. Relation Classification

For models augmented by knowledge graph embeddings, we connect the average encoding of each entity, $x^{word_{h|t}}$, to its knowledge graph embedding, $e^{knowledge_{h|t}}$, to obtain the final input, x, of the RE module.

$$x_{h|t} = [x^{word_{h|t}}, e^{knowledge_{h|t}}] \tag{7}$$

For detailed information about the relationship classification module, we refer to the original works [26,27,31], our relationship classification method adopts the deep biaffine attention mechanism [31]. To capture the directionality of the relations, we use two single-layer feed-forward networks to project pairs of entities into head and tail vector representations, based on their roles as head or tail arguments of the relation.

$$x^h = FFNN_{head}x_h \tag{8}$$

$$x^t = FFNN_{tail}x_t \tag{9}$$

The NEG class is assigned to an entity pair if there is no relation between them, or if the predicted entities are incorrect. Next, we utilize the biaffine attention operator on every head–tail candidate pair, $(x_j^h, x_k^t), j \neq i$:

$$\mathcal{B}(x_1, x_2) := x_1^T U x_2 + W(x_1||x_2) + b \tag{10}$$

$$x_i^{RE} = \mathcal{B}(x_j^{head}, x_k^{tail}) \tag{11}$$

where we indicate with $||$ the column-wise concatenation. To assign scores to each relation class, a softmax activation layer is utilized. The cross-entropy function is employed as the RE loss to compare the predicted scores with the ground truth labels, which are represented through infomative embeddings.

$$\mathcal{L} = \sum_{i=1}^{M} \frac{expr_i}{expx_i^{RE}} \tag{12}$$

5. Experiments

We conduct extensive experiments in standard- and low-resource settings, respectively.

5.1. Datasets

To assess the benefits of incorporating informative labels into sentential RE models, we examined two widely used RE datasets: the Wikidata corpus [9] and the NYT corpus [32].

(i) The Wikidata dataset is generated through distant supervision, connecting the Wikipedia English corpus to Wikidata, and contains sentences with multiple relationships. This extensive dataset includes 353 different relations, with 372,059 training sentences and 360,334 test sentences. (ii) The New York Times Freebase dataset is annotated by linking New York Times articles with Freebase KGs and includes 53 relations (including "NA", indicating no relationship). The training set and test set consist of 455,771 sentences and 172,448 sentences, respectively. To enhance the dataset, we introduce our proposed context.

To evaluate our proposed method in a low-resource setting, we used the FewRel dataset [33,34], which is constructed by remotely aligning Wikidata triples with Wikipedia articles. The FewRel dataset consists of 100 relation types, with 700 instances of each type. The standard FewRel setting adopts 64/16/20 splits for the training/validation/test sets, where the training and validation sets are publicly accessible and the test set is not. We followed the standard Flex benchmark setup and split the training and validation sets of FewRel into a training set of 65 relations, a validation set of 5 relations, and a test set of 10 relations. The test tasks were sampled and processed using the official Flex toolkit.

5.2. Implementation Details

The pre-training knowledge base data, consisting of 25,933,196 sentences, was obtained from Wikipedia Dumps [1]. Wikidata5m [35] integrated the Wikidata knowledge graph and Wikipedia pages, and was utilized as the knowledge base, containing 3,085,345 entities and 822 relation types. To acquire pre-trained graph embeddings for entities and relations, we utilized the method proposed by Adam et al. [30], which is a distributed and memory-efficient implementation capable of handling extensive graphs. This approach has been noted to be as effective as the original implementation of the most advanced models mentioned previously. The pre-trained embeddings we use make use of the TransE [23] approach with an embedding dimension of size 200.

To ensure a fair comparison between models with and without graph embeddings, we excluded sentences with unavailable embeddings from the training phase of the base model. We retained all sentences in the test set, regardless of whether embeddings were available. If an entity had no embedding available, we replaced its graph embedding with a zero tensor.

In this paper, we trained the model for each dataset using a distinct random initialization, repeated ten times both with and without the inclusion of pre-trained graph embeddings. As a result, we trained a total of twenty models for each experiment. It should be noted that throughout the experiments, we consistently employed the AdamW optimizer with a learning rate of 2×10^{-5}. The Pytorch Library [36] was utilized for the implementation of both the optimizer and the overall model.

To evaluate sentential RE, we compare the performance using precision, recall, and the F1 score, both for each single relation class, and on average with micro- and macro-averaging. For zero-shot and few-shot RE tasks, we follow the FLEX benchmark and report the accuracy, confidence interval, and standard deviation correspondingly. All these results reported are from the official Flex toolkits.

5.3. Baselines

For sentence relation extraction (RE), we use three knowledge-augmented RE models as baselines: RECON [5] uses graph neural networks to learn representations of sentences and facts stored in a knowledge graph (KG) to automatically identify relationships in sentences. KGpool [6] dynamically expands the context of a sentence with extra facts from a KG to address the sparsity of sentence RE. PKG [8] combines pretrained knowledge base graph embeddings with transformer-based language models to improve performance on the sentential RE task.

For zero-shot and few-shot RE tasks, we compare our model with UniFew [37], a unified few-shot learning model based on T5 [38]. UniFew converts each few-shot classification task into a machine-reading comprehension format and predicts the results through generation. Pre-trained on large-scale MRC data, UniFew achieves strong performance on both zero- and few-shot RE tasks.

5.4. Overall Performance

5.4.1. Sentential RE

According to the results presented in Table 2, KLA-SRE outperforms both dynamically and statically knowledge integration methods in full-data scenarios. Specifically, KLA-SRE outperforms the PKG_{ge} method by 1.76% on Micro F1 and Macro F1 and 5.45% on NYT, indicating the usefulness of the topological knowledge and description semantics contained in the labels. The topological knowledge and description semantics contained in the labels are useful for classification tasks, thus improving the performance of KLA-SRE.

Our proposed method is outperformed by the current state-of-the-art model KG-Pool in terms of the micro-average metric on the Wikidata dataset, despite the significant improvement in performance achieved by adding knowledgeable labels. However, on the Wikidata dataset, our model (roberta-base) performs similarly to the current state-of-the-art model RECON, showing superior macro-average scores in terms of recall (R) and F1. On

the NYT dataset, our model performs well, improving by 1.65% and 2.57% in micro- and macro-F1, respectively.

Table 2. Table 1 summarizes the average precision (P), recall (R), and F1 scores for two standard datasets obtained from our experiments. The subscript *ge* indicates the model with graph embeddings, and the subscript 1*e* indicates the model trained for one epoch in the Wikidata experiment. We include the results of other studies using the same dataset where possible; otherwise, the spaces are left empty (-). For the New York Times dataset, we only include the single reported F1 score.

			Micro			Macro		
			P	**R**	**F1**	**P**	**R**	**F1**
Wikidata	PKG [8]	RECON [5]	87.24	87.23	87.23	63.59	33.91	44.23
		KG-Pool [6]	88.60	88.59	88.60	-	-	-
		bert-base	85.43	78.37	81.74	51.42	37.24	40.02
		roberta-base	85.50	80.30	82.81	49.33	36.77	40.02
		roberta − $base_{1e}$	83.71	83.01	83.35	46.09	34.52	36.38
	PKG_{ge}	bert-base	88.34	79.19	83.51	55.72	39.62	43.24
		roberta-base	87.66	82.07	84.77	54.42	41.64	44.33
		roberta − $base_{1e}$	86.83	83.93	85.36	50.88	38.52	40.57
	KLA-SRE	bert-base	88.56	82.36	85.35	56.52	40.73	47.34
		roberta-base	88.72	83.93	86.26	56.74	41.23	47.76
		roberta − $base_{1e}$	87.49	86.75	87.12	51.37	40.04	45.02
NYT	PKG	bert-base	47.13	75.57	57.98	28.35	45.27	33.05
	PKG_{ge}	bert-base	51.13	76.46	61.24	31.66	47.31	36.20
	KLA-SRE	bert-base	52.71	77.94	62.89	32.22	48.67	38.77

5.4.2. Low-Shot RE

When compared to the robust baseline method Unifew, the KLA-SRE method achieves similar performance enhancements, as shown in Table 3. KLA-SRE outperforms the baseline methods by at least 1.5% in zero-shot and few-shot RE tasks, exhibiting narrower standard deviations and shallower confidence intervals than Unifew, indicating that its predictions are more consistent across various tasks.

Table 3. The experimental results of the FewRel validation and test sets using zero-shot and few-shot relation extraction are presented in this table.

Model	**Zero-Shot**		**Few-Shot**		**Avg.**
	Acc. $\pm$ ***ci.***	**Std.**	***Acc.*** $\pm$ ***ci.***	**Std.**	
Unifew	52.5 ± 2.0	9.7	79.2 ± 1.5	7.5	65.9
Unifew-meta	79.4 ± 1.9	9.2	87.2 ± 1.2	5.7	83.3
KLA-SRE	53.7 ± 1.7	8.6	81.1 ± 1.1	5.3	84.4

The above two aspects show that explicit knowledge of relation labels is helpful with the pre-trained model, which can reduce the search space for target labels in the early stages of model training. Moreover, Unifew-meta achieves state-of-the-art results for all low-shot RE tasks through supervised meta-training, possibly due to the prevalence of negative instances in the test task. Nonetheless, meta-training programs mitigate this issue.

5.5. Ablation Studies

In this subsection, we study following research questions: "RQ1: How effective is KLA-SRE in capturing the knowledge of relation labels contained in KG structure and description for the sentential RE?" This research question is further divided into two sub-research questions: RQ1.1: what is the useful contribution of each relation attribute context (description semantics, topology information) for sentential RE? RQ1.2: Is the addition of the KG information of relation labels statistically significant?

Table 4 summarizes the performance of KLA-SRE and its configurations. The performance of `w/o relation Demb` outperforms that of `w/o relation Gemb`, indicating that knowledge tags reduce the amount of topological information required to describe the semantic contribution. This reported study on the knowledgeable label-aware relation classification module's effectiveness answers our first sub-research question (RQ1.1). `w/o relation Gemb and entity Gemb` continues to outperform `w/o relation Gemb and relation Demb`. This shows that training in a separate vector space still helps to learn more relationship labels in the relationship classification task. Expressive embeddings and the appropriate leveraging of knowledge stored in knowledge graphs for fitting tasks can resolve representation inconsistencies between sentences and relation labels. The observations can be explained as follows: replacing one hot vector of relationship labels with contextual embeddings from knowledge graph helps the model to correctly predict the relationship class of the entities in the sentence. This positive result verifies the effectiveness of knowledge label incorporation and answers research question RQ1.2.

Table 4. Ablation study on validation set of NYT. w/o, Gemb, and Demb are the abbreviations of without, graph embeddings, and description embeddings.

Model	P	R	F1
KLA-SRE	52.71	77.94	62.89
w/o relation Demb	52.06	76.67	62.01
w/o relation Gemb	51.56	75.83	61.38
w/o relation Gemb and relation Demb	49.26	73.11	58.86
w/o relation Gemb and entity Gemb	50.34	74.25	60.00
w/o relation Demb and entity Gemb	50.52	74.10	60.08

From Table 5, the accuracy decreases when we concatenate random embeddings of the knowledge graph and descriptions instead of utilizing pretrained graph embeddings and on-task description embeddings, indicating the validity of the information contained in the knowledge labels. We further utilize the random embedding of knowledge graphs and descriptions, respectively. The results show that both the relationship description and topology help extract more informative context and beneficial clues for entity embeddings for classification, with the latter being more advantageous. Using meaningful description embeddings alone only slightly improves the original network without any embeddings, suggesting that the topology may provide more instructional cues. This is because topology is more consistent with the characteristics of a particular relationship. Replacing pretrained graph embeddings and description embeddings via different representation methods can achieve comparable performance. All these results answer RQ1.

Table 5. Ablation study on the FewRel dataset. Gemb and Demb are the abbreviations of graph embeddings and description embeddings.

Model	$Acc. \pm ci.$	Std.
KLA-SRE	81.1 ± 1.1	5.3
- Random Embeddings of Gemb and Demb	71.1 ± 1.0	6.1
- Random Embeddings of Gemb	75.4 ± 1.2	7.3
- Random Embeddings of Demb	79.2 ± 1.6	7.3
- RESCAL [39] Gemb	80.5 ± 1.3	6.7
- DisMult [40] Gemb	80.9 ± 1.6	6.9
- Glove Demb	79.8 ± 1.7	7.2
- no Gemb and Demb	69.4 ± 1.8	7.7

6. Conclusions and Future Work

We propose a neural architecture that leverages the KG topology information and description semantics of relation labels for sentential relation extraction. Our model significantly outperforms the baselines' high-resource settings and performs on par in the low-resource setting. We performed extensive experiments to show that the knowledgeable label-aware relation classification module incorporates strong prior knowledge from KGs. We demonstrate that the consistency of relational label representations and sentence representations is critical to retrieving prior knowledge.

Based on the assessment of ornamental value and the insights gained in this article, we point out future research directions: (1) Researchers with access to the industrial research ecosystem can investigate how KLA-SRE and other sentential REs can be evaluated for industrial applications, and (2) the impact of retrieved KG data quality on sentential RE performance is a feasible next step for future research.

Author Contributions: Conceptualization, B.N. and Y.S.; writing—review and editing, B.N. All authors have read and agreed to the published version of the manuscript.

Funding: Supported by the Fundamental Research Funds for the Provincial Universities of Zhejiang (GrantNum:GK229909299001-022).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: We declare that we have no financial and personal relationships with other people or organizations that can appropriately influence our work, and there is no professional or other personal interest of any nature or kind in my product, service, and/or company that could be constructed as influencing the position presented in, or the review of, the manuscript entitled.

References

1. Vrandečić, D. Wikidata: A new platform for collaborative data collection. In Proceedings of the Web Conference , Lyon, France, 16–20 April 2012; pp. 1063–1064.
2. Suchanek, F.M.; Kasneci, G.; Weikum, G. Yago: A Core of Semantic Knowledge Unifying WordNet and Wikipedia. In Proceedings of the 16th International Conference on World Wide Web, Banff, AB, Canada, 8–12 May 2007; pp. 697–697.
3. Bodenreider, O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res.* **2004**, *32*, D267–D270. [CrossRef]
4. Mintz, M.D.; Bills, S.; Snow, R.; Jurafsky, D. Distant supervision for relation extraction without labeled data. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, Suntec, Singapore, 2–7 August 2009; pp. 1003–1011.
5. Bastos, A.; Nadgeri, A.; Singh, K.; Mulang, I.O.; Shekarpour, S.; Hoffart, J. RECON: Relation Extraction using Knowledge Graph Context in a Graph Neural Network. In Proceedings of the Web Conference, Ljubljana, Slovenia, 19–23 April 2021; pp. 1673–1685.
6. Nadgeri, A.; Bastos, A.; Singh, K.; Mulang, I.O.; Hoffart, J.; Shekarpour, S.; Saraswat, V.A. KGPool: Dynamic Knowledge Graph Context Selection for Relation Extraction. In Proceedings of the Findings of the Association for Computational Linguistics: ACL-IJCNLP, Virtual Event, 1–6 August 2021; pp. 535–548.
7. Vashishth, S.; Joshi, R.; Prayaga, S.S.; Bhattacharyya, C.; Talukdar, P.P. RESIDE: Improving Distantly-Supervised Neural Relation Extraction using Side Information. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 1257–1266.
8. Papaluca, A.; Krefl, D.; Suominen, H.; Lenskiy, A. Pretrained Knowledge Base Embeddings for improved Sentential Relation Extraction. In Proceedings of the Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, Dublin, Ireland, 22–27 May 2022; pp. 373–382.
9. Sorokin, D.; Gurevych, I. Context-Aware representations for knowledge base relation extraction. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, EMNLP, Copenhagen, Denmark, 9–11 September 2017; pp. 1784–1789.
10. Zhou, W.; Chen, M. An Improved Baseline for Sentence-level Relation Extraction. In Proceedings of the Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing, Online, 20–23 November 2022; pp. 161–168.

11. Xu, B.; Liu, N.; Cheng, L.; Huang, S.; Wei, S.; Du, M.; Song, H.; Wang, H. Knowledge Graph Enhanced Sentential Relation Extraction via Dual Heterogeneous Graph Context Selection . In Proceedings of the International Joint Conference on Neural Networks, Gold Coast, Australia, 18–23 June 2023; pp. 1–7.
12. Soares, L.B.; FitzGerald, N.; Ling, J.; Kwiatkowski, T. Matching the blanks: Distributional similarity for relation learning. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 2895–2905.
13. Zhang, Z.; Han, X.; Liu, Z.; Jiang, X.; Sun, M.; Liu, Q. ERNIE: Enhanced language representation with informative entities. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 1441–1451.
14. Peters, M.E.; Neumann, M.; Logan, R.; Schwartz, R.; Joshi, V.; Singh, S.; Smith, N.A. Knowledge enhanced contextual word representations. In Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, Hong Kong, China, 3–7 November 2019; pp. 43–54.
15. Weston, J.; Bordes, A.; Yakhnenko, O.; Usunier, N. Connecting Language and Knowledge Bases with Embedding Models for Relation Extraction. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Grand Hyatt Seattle, Seattle, WA, USA, 18–21 October 2013; pp. 1366–1371.
16. Han, X.; Liu, Z.; Sun, M. Neural Knowledge Acquisition via Mutual Attention Between Knowledge Graph and Text. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; pp. 4832–4839.
17. Zhang, N.; Deng, S.; Sun, Z.; Wang, G.; Chen, X.; Zhang, W.; Chen, H. Long-tail Relation Extraction via Knowledge Graph Embeddings and Graph Convolution Networks. In Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; pp. 3016–3025.
18. Hu, L.; Zhang, L.; Shi, C.; Nie, L.; Guan, W.; Yang, C. Improving Distantly-Supervised Relation Extraction with Joint Label Embedding. In Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, Hong Kong, China, 3–7 November 2019; pp. 3821–3829.
19. Dai, Q.; Inoue, N.; Reisert, P.; Takahashi, R.; Inui, K. Incorporating Chains of Reasoning over Knowledge Graph for Distantly Supervised Biomedical Knowledge Acquisition. In Proceedings of the Pacific Asia Conference on Language, Information and Computation, Hakodate, Japan, 13–15 September 2019.
20. Hu, Z.; Cao, Y.; Huang, L.; Chua, T. How Knowledge Graph and Attention Help? A Qualitative Analysis into Bag-level Relation Extraction. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Virtual Event, 1–6 August 2021.
21. Angeli, G.; Johnson, M.; Manning, C.D. Leveraging Linguistic Structure For Open Domain Information Extraction. In Proceedings of the Annual Meeting of the Association for Computational Linguistics and the International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Beijing, China, 26–31 July 2015; pp. 344–354.
22. Schuster, M.; Paliwal, K.K. Bidirectional recurrent neural networks. *IEEE Trans. Signal Process.* **1997**, *45*, 2673–2681. [CrossRef]
23. Bordes, A.; Usunier, N.; García-Durán, A.; Weston, J.; Yakhnenko, O. Translating Embeddings for Modeling Multi-relational Data. In Proceedings of the International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–8 December 2013; pp. 2787–2795.
24. Zhu, H.; Lin, Y.; Liu, Z.; Fu, J.; Chua, T.S.; Sun, M. Graph Neural Networks with Generated Parameters for Relation Extraction. *arXiv* **2019**, arXiv:1902.00756.
25. Kipf, T.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv* **2016**, arXiv:1609.02907.
26. Nguyen, D.Q.; Verspoor, K.M. End-to-end neural relation extraction using deep biaffine attention. In Proceedings of the European Conference on Information Retrieval, Grenoble, France, 26–29 March 2018; pp. 729–738.
27. Giorgi, J.; Wang, X.; Sahar, N.; Shin, W.Y.; Bader, G.D.; Wang, B. End-to-end Named Entity Recognition and Relation Extraction using Pre-trained Language Models. *arXiv* **2019**, arXiv:1912.13415.
28. Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; et al. HuggingFace's Transformers: State-of-the-art Natural Language Processing. *arXiv* **2019**, arXiv:1910.03771.
29. Araci, D. FinBERT: Financial Sentiment Analysis with Pre-trained Language Models. *arXiv* **2019**, arXiv:1908.10063.
30. Lerer, A.; Wu, L.Y.; Shen, J.; Lacroix, T.; Wehrstedt, L.; Bose, A.; Peysakhovich, A. PyTorch-BigGraph: A Large-scale Graph Embedding System. *arXiv* **2019**, arXiv:1903.12287.
31. Dozat, T.; Manning, C.D. Deep Biaffine Attention for Neural Dependency Parsing. *arXiv* **2016**, arXiv:1611.01734.
32. Riedel, S.; Yao, L.; McCallum, A. Modeling Relations and Their Mentions without Labeled Text. In Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases, MIT Stata Center, MA, USA, 9–11 October 2010.
33. Han, X.; Zhu, H.; Yu, P.; Wang, Z.; Yao, Y.; Liu, Z.; Sun, M. FewRel: A Large-Scale Supervised Few-Shot Relation Classification Dataset with State-of-the-Art Evaluation. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 4803–4809.
34. Gao, T.; Han, X.; Zhu, H.; Liu, Z.; Li, P.; Sun, M.; Zhou, J. FewRel 2.0: Towards More Challenging Few-Shot Relation Classification. In Proceedings of the Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 6251–6256.

35. Wang, X.; Gao, T.; Zhu, Z.; Liu, Z.; Li, J.-Z.; Tang, J. KEPLER: A Unified Model for Knowledge Embedding and Pre-trained Language Representation. *Trans. Assoc. Comput. Linguist.* **2019**, *9*, 176–194. [CrossRef]
36. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv* **2019**, arXiv:1912.01703.
37. Bragg, J.; Cohan, A.; Lo, K.; Beltagy, I. FLEX: Unifying Evaluation for Few-Shot NLP. In Proceedings of the Neural Information Processing Systems, virtual, 6–14 December 2021; pp. 15787–15800.
38. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* **2020**, *21* , 1–67.
39. Nickel, M.; Tresp, V.; Kriegel, H.P. A Three-Way Model for Collective Learning on Multi-Relational Data. In Proceedings of the 28th International Conference on International Conference on Machine Learning, Bellevue, WA, USA, 28 June–2 July 2011; pp. 809–816.
40. Yang, B.; Yih, W.T.; He, X.; Gao, J.; Deng, L. Embedding Entities and Relations for Learning and Inference in Knowledge Bases. *arXiv* **2014**, arXiv:1412.6575.

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.