Logic-guided Semantic Representation Learning for Zero-Shot Relation Classification

Juan Li[1,2]*, **Ruoxu Wang**[1,2]*, **Ningyu Zhang**[1,2]*†, **Wen Zhang**[1,2],
Fan Yang[1,2], **Huajun Chen**[1,2] †
[1] Zhejiang University
[2] AZFT Joint Lab for Knowledge Engine
{lijuan18,ruoxuwang,zhangningyu,wenzhang2015,21821249,huajunsir}@zju.edu.cn

Abstract

Relation classification aims to extract semantic relations between entity pairs from the sentences. However, most existing methods can only identify seen relation classes that occurred during training. To recognize unseen relations at test time, we explore the problem of zero-shot relation classification. Previous work regards the problem as reading comprehension or textual entailment, which have to rely on artificial descriptive information to improve the understandability of relation types. Thus, rich semantic knowledge of the relation labels is ignored. In this paper, we propose a novel logic-guided semantic representation learning model for zero-shot relation classification. Our approach builds connections between seen and unseen relations via implicit and explicit semantic representations with knowledge graph embeddings and logic rules. Extensive experimental results demonstrate that our method can generalize to unseen relation types and achieve promising improvements.

1 Introduction

Relation Classification (RC) is an important task in information extraction, aiming to extract the relation between two given entities based on their related context. RC has attracted increasing attention due to its broad applications in many downstream tasks, such as knowledge base construction (Luan et al., 2018) and question answering (Yu et al., 2017).

Conventional supervised RC approaches can not satisfy the practical needs of the relation classification. In the real world, there exist massive amounts of fine-grained relations. And, the labeled relation types are limited, and each type usually has a certain number of labeled samples. Naturally, it is prohibitive to generalize to new (unseen) relations (i.e., the model will fail when predicting a type with no training examples). For example, in Figure 1, *basin_country* is an unseen relation type with no labeled sentence in the training stage. To this end, it is urgent for models to be able to extract relations in a zero-shot scenario.

Previous zero-shot relation classification (ZSRC) approaches leverage transfer learning procedures by reading comprehension (Levy et al., 2017), textual entailment (Obamuyide and Vlachos, 2018), and so on. However, those methods have to rely on artificial descriptive information to improve the understandability of relation types. Inspired by the zero-shot learning in computer vision (Palatucci et al., 2009), it is natural to learn a mapping from the feature space of input samples to the semantic space such as class labels through a projection function. The hypothesis is to build the **semantic connections** between seen and unseen relations. Conventional approaches usually leverage word embeddings (Mikolov et al., 2013) of labels as a common semantic space. We argue that for relation classification, rich semantic knowledge is neglected in the relation labels space:

Implicit Semantic Connection with Knowledge Graph Embedding. Previous studies (Yang et al., 2014) have shown that the Knowledge Graph Embeddings (KGEs) of semantically similar relations are
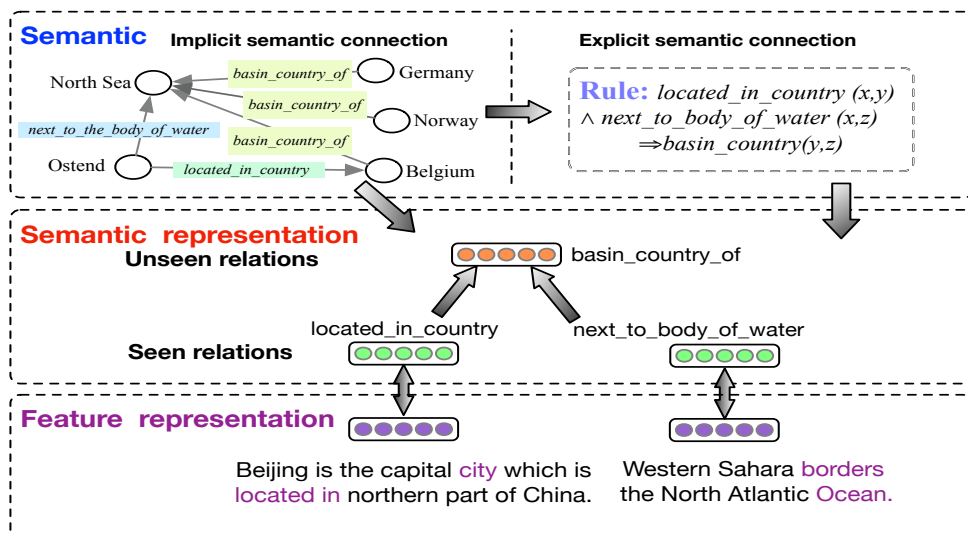

Figure 1: Knowledge graph embedding and rule learning for zero-shot relation classification.

located near each other in the latent space. For instance, the relation *place_lived* and *nationality* are more relevant, whereas the relation *profession* has less correlation with the former two relations. Thus, it is natural to leverage this knowledge from KGs to build connections between seen and unseen relations.

Explicit Semantic Connection with Rule Learning. We human can easily recognize unseen relations via symbolic reasoning. As the example shown in Figure 1, with the rule that *basin_country_of(y,z)* can be deduced if *located_in_country(x,y)* and *next_to_body_of_water(x,z)*, we can recognize the unseen relation *basin_country_of* based on seen relations *located_in_country* and *next_to_body_of_water*. To this end, it is intuitive to infuse rule knowledge to bridge the connections between seen and zero-shot relations.

Motivated by this, we take the first step to propose a novel approach, namely, **L**ogic-guided **S**emantic **R**epresentation **L**earning (**LSRL**) for zero-shot relation classification. To begin with, we propose to utilize pre-trained knowledge graph embedding such as TransE (Bordes et al., 2013) to build the *implicit* semantic connection. KGE embeds entities and relations into a continuous semantic vector space and can capture semantic connections between relations in semantic space. Further, we leverage logic rules mined via AMIE (Galárraga et al., 2013) from the knowledge graph and introduce rule-guided representation learning to obtain *explicit* semantic connection. It should be noted that our approach is model-agnostic, and therefore orthogonal to existing approaches. We integrate our approach with two well-known zero-shot methods, namely DeViSE (Frome et al., 2013) and ConSE (Norouzi et al., 2013). Extensive experimental results demonstrate the efficacy of our approach.

The main contributions of this work are as follows:

- We introduce implicit semantic connection with knowledge graph embedding and explicit semantic connection with rule learning for zero-shot relation classification.

- We propose a novel rule-guided semantic representation learning to build connections between the seen and unseen relations. Our work is model-agnostic and can be plugged into different kinds of zero-shot learning approaches.

- Extensive experimental results show the efficacy of our approach and also reveals the usefulness of knowledge graph embedding and rule learning.

2 Related Work

Relation Classification. Relation classification (RC) has been firstly proposed in MUC 1998, which aims to predict the relation between two entities by a specific context. Many mature models have been developed to figure out this problem, including traditional methods like (GuoDong et al., 2005), deep neural networks approach like (Zeng et al., 2015; Zhang et al., 2018; Zhang et al., 2019a; Deng et al., 2020a; Deng et al., 2020b; Zhang et al., 2020a; Zhang et al., 2020b; Yu et al., 2020; Wang et al., 2020), and some joint models like (Zheng et al., 2017; Ye et al., 2020). However, those methods are all

supervised approaches which can only infer relations existing in the train set but are incapable of making predictions for newly-add relations.

Zero-shot relation classification (ZSRC) was first proposed by (Levy et al., 2017), which is able to extract new relations by reducing relation classification to answering simple reading comprehension questions. Lately, (Obamuyide and Vlachos, 2018) formulates relation extraction as a textual entailment problem and considers the input instance and relation description as the premise and hypothesis. However, these approaches require human annotators to construct questions or write descriptions for relations, which is labor-intensive. On the contrary, our zero-shot relation classification approach does not need any human involvement and can be integrated into most existing RC models.

Zero-shot Learning In the computer vision field, zero-shot learning (ZSL) has attracted a lot of attention. The key that underpins ZSL in image recognition is to exploit the shared semantic representations between seen and unseen classes and transfer them to the visual representations of samples. (Frome et al., 2013) proposes a ZSL model called DeViSE to learn a linear mapping between image features and semantic space using an efficient ranking loss formulation. (Norouzi et al., 2013) proposes ConSE, which first predicts seen class posteriors, then projects image features into the word2vec space by considering the convex combination of top T most possible seen classes. The semantic representation of those approaches is learned by certain auxiliary information attached to the class labels, such as attribute description (Jayaraman and Grauman, 2014; Farhadi et al., 2009) and embedding representation (Romera-Paredes and Torr, 2015; Akata et al., 2016). Different from the zero-shot approaches in computer vision, we construct the semantic space by considering information from the knowledge graph rather than word embedding or attribute.

There are also some ZSL applications in NLP, such as event extraction (Huang et al., 2017), entity-typing (Zhou et al., 2018), cross-lingual entity linking (Rijhwani et al., 2018), text classification (Pushp and Srivastava, 2017) and cold-start recommendation (Li et al., 2019), as well as in KG such as link prediction(Qin et al., 2020).

Knowledge Graph Embedding. In recent years, various KG embedding methods, including translation-based, semantic matching and neural network methods, have been devised to learn vector representations for entities and relations of a KG. Translation-based models (Bordes et al., 2013; Ji et al., 2015; Lin et al., 2015b) use distance-based scoring functions to assess the plausibility of a triple. For example, in TransE (Bordes et al., 2013), the score function is $f_r(h,t) = ||h + r - t||_{l_{1/2}}^2$. Semantic matching models (Yang et al., 2014; Nickel et al., 2016; Liu et al., 2017) employ similarity-based scoring functions to compute the energy of relational triples, where the scoring function of the representative model DistMult (Yang et al., 2014) is $f_r(h,t) = h^\top diag(M_r)t$. Neural network models learn to express entities and relations through neural networks, such as CNN-based methods (Dettmers et al., 2018) and GNN-based methods (Schlichtkrull et al., 2018). We utilize KG embedding models to learn the representations of relations instead of word embeddings so that the representations of relations are only related to the structure of a KG but not relations' name. Meanwhile, connections of relations are harvested.

Rule Learning. Rules over a KG can capture connections between relations, and a variety of methods for rule learning have been studied, such as Inductive Logic Programming(ILP) algorithms, rule mining methods, and embedding-based methods. ILP is formalized by first-order logic and has strong representation powers, but does not scale to large datasets (Sadeghian et al., 2019). To address this, several efficient rule miners for KGs have been developed, such as RDF2rules (Wang and Li, 2015), ScaleKB (Chen et al., 2016) and AMIE+ (Galárraga et al., 2015). In addition, embedding-based rule learning methods have gained attention. RLvLR (Omran et al., 2018) utilizes embeddings to guide rule extraction and reduce the search space. DistMult (Yang et al., 2014) utilizes learned embeddings of entities and relations to extract logical rules. And (Ho et al., 2018) introduces a framework for rule learning guided by external sources. We adopt the widely used rule mining method proposed in (Galárraga et al., 2013) to extract rules from KG in this paper. To the best of our knowledge, we are the first approach to address zero-shot relation classification with the assistance of logical rules from KG.

3 Methodology

3.1 Preliminaries

We start by defining some notations and terms. $\mathcal{R}^S$ denotes the set of seen relations during training and $\mathcal{R}^U$ denotes the set of unseen relations for testing, and $\mathcal{R}^S \bigcap \mathcal{R}^U = \phi$. $\mathcal{D}^{tr} = \{(s_i, h_i, t_i, y_i), i = 1, ..., N_s\}$ denotes training dataset, where s_i represents one sentence and (h_i, t_i) is an entity pair mentioned in s_i. N_s is the number of seen relations and $y_i \in \mathcal{R}^S$ denotes the relation of the entity pair. Analogously, $\mathcal{D}^{ts} = \{(s_j, h_j, t_j, y_j), j = 1, ..., N_u\}$ denotes the test dataset, where N_u is the number of unseen relation and $y_j \in \mathcal{R}^U$ denotes the relation of h_j and t_j in the sentence s_j.
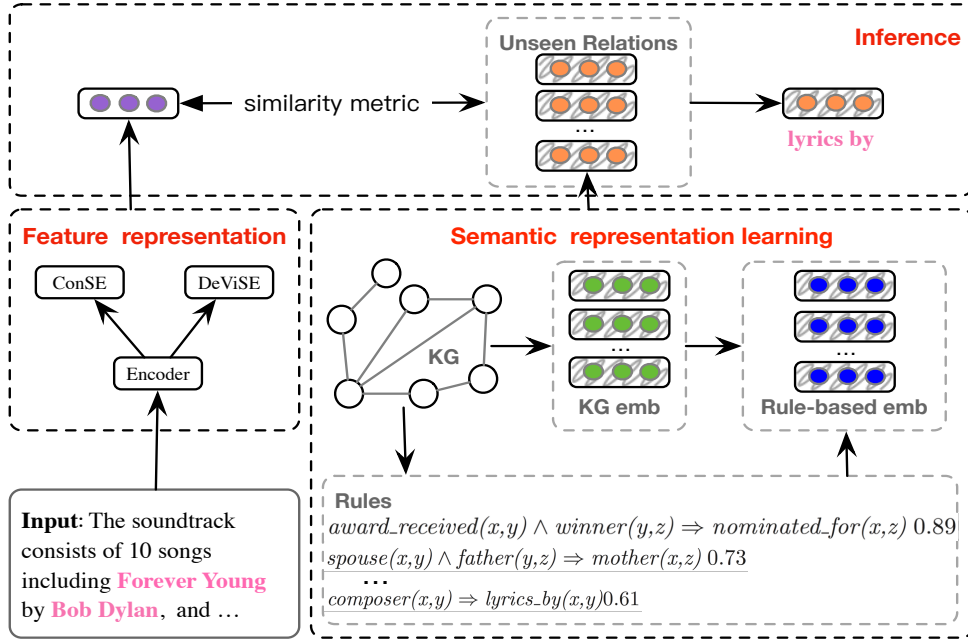

Figure 2: The architecture of Logic-guided Semantic Representation Learning model.

The overall framework is illustrated in Figure 2, which is composed of three modules:

Feature Representation (§3.2) encodes the input sentence into the feature space, which is aimed at capturing syntax features of sentences.

Semantic Representation Learning (§3.3) maps relation types into a semantic space and builds up the connections between seen and unseen relations. Specifically, we propose logic-guided semantic representation learning with knowledge graph embedding and rule learning.

Inference (§3.4) predicts relation types via computing similarity between feature representation of current input sentence and semantic representations for all unseen relations. We infer the unseen relation with the label that is most similar in the semantic space.

3.2 Feature Representation

The input of feature representation is a sentence, and the output is its vector representation. Firstly, we use the Piecewise Convolutional Neural Networks(PCNNs) (Zeng et al., 2015) model to encode input instance, and then use two types of projection functions including *DeViSE* and *ConSE* to get the final feature representation of the input instance.

PCNNs has been proven to be effective in RC. It inputs the concatenation of word embedding and position embedding into a Convolution Neural Network(CNN) to obtain the hidden layer representation h. Then, h is divided into three parts based on the two entities' positions, and max pooling is perfomed on each part to obtain (pl_1, pl_2, pl_3). Final feature encoding of the input sentence $f = [pl_1; pl_2; pl_3]$ is the concatenation of the three pooling segments. We denote the process as below:

$$f = PCNN(x_1, ..., x_n) \tag{1}$$

DeViSE formulates ZSL as a regression problem and learns a linear function to project the input representation to target semantic space:

$$g = W * f + b \tag{2}$$

ConSE maps input representation into target semantic space via convex combination. It trains a classifier C on training dataset D^{tr} and obtains top T probable seen relation types R_t^S together with their probability p_t. $E(R_t^S)$ is the embedding of R_t^S, and then weighted sum on $E(R_t^S)$ is regarded as the feature representation of inputs. The process can be formulated as follows:

$$R_t^S, p_t, E(R_t^S) = C(f), t = 1, ..., T \tag{3}$$

$$g = \sum_{t=1}^{T} p_t * E(R_t^S) \tag{4}$$

3.3 Semantic Representation Learning

Semantic representation builds connections between unseen and seen relations in ZSRC via external resources. We describe the following three kinds of embedding representations in a semantic space.

Word Embedding denoted as E_{wd} is the commonly used method. However, this way faces challenges as analyzed in the introduction. In order to capture the rich explicit or implicit semantic connection between relations, two forms of embedding methods based on KG are introduced.

KG Embedding embeds relations and entities into latent low-dimensional continuous-space vectors, denoted as E_{kg}. In KG embedding methods, the score of a triple (h,r,t) can be calculated via head entity embedding E(h), relation embedding E(r), and tail embedding E(t). Different methods follow different assumptions. For example, the typical method TransE (Bordes et al., 2013) supposes a translation law in the semantic space where $E(h) + E(r) = E(t)$ for positive triples in a KG. Therefore, relation embedding from KG embedding methods is related to triples it belongs to and is not affected by what words it contains. While sometimes the word contained in the relation string also reveals the semantic of this relation, word embedding and KG embedding can complement each other at that time. Hence, we also consider combining them together through a linear transformation, **KG+Word embedding** is defined as:

$$E_{kw} = W_2 * ([E_{kg}; E_{wd}] + b_2) \tag{5}$$

where $[x; y]$ means concatenation of x and y.

Rule-guided Embedding represents rules in vector space instead of symbolization, denoted as E_{rl}. In a knowledge graph, logic rules show the connections between relations. They are in the form of *body* $\Rightarrow$ *head*, where *head* is a binary atom and *body* is a conjunction of binary and unary atoms, such as rule *spouse(x,y)* $\wedge$ *father(y,z)* $\Rightarrow$ *mother(x,z)*, and the number of atoms in *body* is the length of the corresponding rule. We adopt typical rule mining methods such as AMIE (Galárraga et al., 2013) to generate rules from structural KGs. In addition to rules, AMIE also produces the PCA confidence $conf$ to filter out rules. Inspired by (Zhang et al., 2019b), we apply an simple but effective embedding-based method to incorporate symbolic rules into semantic space and generate E_{rl}. Taking TransE as an example, it assumes $h + r \approx t$ for a positive triple (h, r, t). According to this assumption, we can get $r_1 + r_2 = r_3$ if rule $r_1(x, y) \wedge r_2(y, z) \Rightarrow r_3(x, z)$ exists as mentioned in (Lin et al., 2015a). Thus if the rule contains an unseen relation, embedding of the unseen relation can be calculated based on other seen relations in this rule. Besides, it is possible that one unseen relation involves multiple rules, for that, we calculate unseen relation's embedding as follows:

$$E_{rl}(R_i^U) = \frac{\sum_{j=1}^{K} conf_j * E_{kg}(Rule_{ij}^U)}{\sum_{j=1}^{K} conf_j} \tag{6}$$

where R_i^U represents the i_{th} unseen relation, $Rule_{ij}^U$ is the j_{th} rule in the set of rules about R_i^U with top K highest PCA confidence score and $conf_j$ represents the PCA confidence of $Rule_{ij}^U$. For example, with two rules about unseen relation r, $R1 : r_A \wedge r_B \Rightarrow r$ and $R2 : r_C \wedge r \Rightarrow r_D$, following TransE's assumption, we calculate embedding of r via $E_{rl}(r) = \frac{conf_1*[E_{kg}(r_A)+E_{kg}(r_B)]+conf_2*[E_{kg}(r_D)-E_{kg}(r_C)]}{conf_1+conf_2}$.

Similar to Word+KG embedding, we also consider **KG+Rule Embedding**, denoted as E_{kr}, as they might also complement each other. E_{kr} can be calculated as follows:

$$E_{kr} = \lambda * E_{rl} + (1 - \lambda) * E_{kg} \tag{7}$$

where λ is a hyperparameter representing the combination weight between KG embedding and rules. It is set as 0.5 in our experiment. Meanwhile, we calculate **Rule+Word Embedding**, denoted as E_{rw}, by replacing E_{kg} in Equation 5 with E_{rl}.

3.4 Inference

During prediction, we compare the similarity between feature representation f of input sentence and the semantic representation of unseen relations as follows:

$$\overline{y_i} = sim(f_{x_i}, E(R_{x_i}^U)) \tag{8}$$

The similarity function $sim()$ can be cosine similarity or Euclidean Distance. Unseen relations with higher similarity with sentence feature representation are more likely to be predicted.

4 Experiment

In experiments, we want to explore: 1) Whether embeddings based on KGs are more useful for the ZSRC task than word embeddings? 2) What is the factor that can strengthen semantic representations in ZSRC? 3) Whether and how can logical knowledge help build better semantic space?

4.1 Datasets

Different from zero-shot learning relation classification dataset of (Levy et al., 2017) and (Obamuyide and Vlachos, 2018), our method considers rules of relations during training rather than question templates or relation descriptions. We construct a new dataset based upon Wikipedia-Wikidata (Sorokin and Gurevych, 2017) relation extraction dataset which contains 353 relations and 856,217 instances. To evaluate the capability of injecting rule logic into the zero-shot prediction models, we ensure that relations have certain connections in our dataset. We firstly cluster all 353 relations based on word embeddings, then divide seen and unseen relations according to the instance number of relations by the given threshold (1200) for one cluster. We drop relations from the cluster where all relations' instance number is less than 500 with the assumption that there is no support from related seen labels. Manual adjustments are further applied to get the final dataset. In this dataset, there are ultimately 100 relations, 70 of which are seen relations and the rest 30 are unseen ones.

4.2 Settings

Modules	Parameters	Settings(Tried)	Settings
	Kernel Size	3,4,5	3
	Position Embedding Size	5,10,15	5
PCNN	Learning Rate	0.01,0.05,0.005	0.01
	Dropout	0,0.1,0.3,0.5	0.5
	Channel Number	150,200,250,300	250
ConSE	T	1,2,3	3
DeViSE	Margin	0.8,1.0,1.2	1.0

Table 1: Parameter Settings

We use Wikipedia documents[1] to train **word embeddings**, where words appear more than ten times are preserved in vocabulary, following a common setting as other works. Word2vec (Mikolov et al., 2013) is applied for word embedding training with window size set as 5. For **KG embeddings**, we use

---

[1] https://dumps.wikimedia.org/enwiki/latest/enwiki-latest-pages-articles.xml.bz2

TransE to train the embedding of entities and relations on Wikidata, which contains about 20,982,733 entities and 594 relations in total. The embedding size is set as 100, the margin as 1.0, and the learning rate as 0.01. For the PCNN layer, we set kernel size as 3, position embedding size as 5, the number of channels as 250, margin as 2.0, learning rate as 0.01, and dropout as 0.5. For ConSE, Top 3 seen classes are chosen for prediction. For DeViSE, the margin is set as 1.0. For rule mining, we set the max length of rules as 2.

4.3 Whether KG-based embeddings are more useful than word embeddings?

	ConSE(Hit@n)			DeViSE(Hit@n)		
	1	2	5	1	2	5
$+E_{wd}$	0.21	0.30	0.43	0.11	0.19	0.39
$+E_{kg}$	0.39	0.53	0.69	0.22	0.38	0.57
$+E_{rl}$	0.40	0.54	0.72	0.23	0.39	0.58
$+E_{kw}$	0.39	0.55	0.72	0.23	**0.40**	**0.59**
$+E_{rw}$	0.40	0.55	0.70	0.23	0.34	0.57
$+E_{kr}$	**0.43**	**0.57**	**0.74**	**0.25**	0.39	**0.59**

Table 2: Performance of DeViSE and ConSE in the case of different embedding methods, including Word(E_{wd}), KG(E_{kg}), Rule(E_{rl}), KG+Word(E_{kw}), Rule+Word(E_{rw}) and KG+Rule(E_{kr}) embeddings.

To compare the effectiveness of KG-based embeddings and word embeddings, we regard methods that use E_{wd} as baselines. Results of two kinds of methods, including KG-based embeddings and a combination of any two embeddings, are listed to explore the usefulness of KGs in the ZSRC task. The experiments also distinguish the results when using two different projection functions in the ZSL problem, i.e., ConSE and DeViSE. During testing, we rank the similarity scores between feature representations of test sentences and all unseen relations' semantic representations based on cosine similarity and get ranks of true labels. Hit@K(K=1, 2, 5) are used as evaluation metrics.

The overall results are shown in Table 2. Under the ConSE structure, methods that incorporate semantic representation based on the knowledge graph significantly outperform word embedding. Specifically, KG embedding gains improvement with 18% and rule embedding with 19% on Hit@1. The performance of word+KG embedding and word+rule-based embedding also improves a lot, and the combination of KG+rule-based embedding achieves the best performance. Thus, we can conclude that *KG-based embeddings are superior to word embedding*.

Additional inspection of the table shows that the results over ConSE are better than DeViSE in all embedding settings. The reason is associated with the difference between the two models. The representation space of ConSE is limited to a combined space consisting of seen classes, while DeViSE enables mapping instance embedding to the whole relation space. Thus the dataset with stronger relevance between relations is more friendly to ConSE, which is the case in our ZSRC dataset.

4.4 What is the factor that can strengthen semantic representations in ZSRC?

We analyze the question via a comparison between word embeddings and KG embeddings. A closer inspection of a specific relation is illustrated in Table 3. For most unseen relations, KG embeddings perform better than word embeddings. Especially for *drafted_by*, *occupant* and *office_contested*, word embeddings predict almost nothing, while KG embeddings achieve 81%, 31% and 26% respectively. The reason may be that word embedding is less than enough to capture complete, accurate, or even logic-level connections between relations. For example, for the relation *drafted_by* which means "allocate certain players to teams in some sports", it is difficult for word embedding to capture its connection to the relations *member_of_sports_team* and *educated_at* because the word *draft* has other senses such as "draft a document" that appears much more commonly in the corpus. By contrast, KG embeddings are trained based upon the entity pairs and relationships existing within a whole knowledge base, thereby capturing

Unseen Relations	F1-score		Top 3 Related Seen Relations	
	$+E_{kg}$	$+E_{wd}$	$+E_{kg}$	$+E_{wd}$
lyrics_by	**0.52**	0.06	performer	influenced_by
			composer	spouse
			cast_member	cast_member
after_a_work_by	**0.51**	0.01	author	named_after
			screenwriter	author
			creator	characters
location_of_formation	**0.46**	0.02	headquarters_location	subclass_of
			location	opposite_of
			capital	part_of
nominated_for	**0.97**	0.56	award_received	award_received
			winner	part_of
			participant_of	member_of
mother	0.40	**0.83**	follows	child
			spouse	spouse
			twinned_administrative_body	father
developer	0.38	**0.49**	publisher	manufacturer
			manufacturer	publisher
			owned_by	owned_by
office_contested	**0.26**	0.00	position_held	
			successful_candidate	———
			applies_to_jurisdiction	
occupant	**0.31**	0.00	owned_by	
			location	———
			headquarters_location	
drafted_by	**0.81**	0.00	member_of_sports_team	
			educated_at	———
			member_of	

Table 3: Results of KG embedding and word embedding on F1 score when using ConSE as projection function. And top 3 most influential seen relations of the corresponding unseen relation are presented.

the more accurate meaning of, and connections between these relations.

Negative examples are also found in Table 3, such as *mother*, for which KG embedding predicts poorly, but word embedding achieves 83%. The reason may be that the number of training triples for *mother* is relatively small in our dataset, leading to poor embeddings. KG embedding suffers from its sparsity problem due to imperfect KG, whereas word embedding excels at capturing contextually similar words such as *mother* to *father* or *spouse*.

These results show that *successfully building accurate or even logical-level connections between seen and unseen relations is an essential factor for zero-shot tasks, and this is why KG embeddings perform better than word embeddings.*

Unseen Relations	F1-score						Related rules w.r.t. unseen relations
	$+E_{wd}$	$+E_{kg}$	$+E_{rl}$	$+E_{kw}$	$+E_{rw}$	$+E_{kr}$	
mother	**0.83**	0.40	0.77	0.53	0.80	0.78	*mother(x,z) ⇐ spouse(x,y) ∧ father(y,z)* *mother(x,y) ⇐ child(y,x)*
lyrics_by	0.06	0.52	0.51	0.49	0.48	**0.52**	*lyrics_by(x,y) ⇐ composer(x,y)*
nominated_for	0.56	**0.97**	0.96	**0.97**	0.96	0.96	*nominated_for(x,z) ⇐ award_received(x,y) ∧ winner(y,z)*
producer	0.41	0.52	**0.55**	0.54	0.52	0.53	*producer(x,y) ⇐ director(x,y)* *producer(x,y) ⇐ screenwriter(x,y)* *producer(x,y) ⇐ cast_member(x,y)*
field_of_work	0.04	0.14	0.29	0.11	0.29	**0.37**	*field_of_work(x,y) ⇐ occupation(x,y)*
connecting_line	0.00	0.10	0.43	0.28	0.42	**0.47**	*connecting_line(x,z) ⇐ adjacent_station(y,x) ∧ part_of(y,z)*
residence	0.01	0.32	0.30	0.30	0.38	**0.39**	*residence(x,y) ⇐ place_of_birth(x,y)* *residence(x,y) ⇐ place_of_death(x,y)*

Table 4: Results of all different embeddings on F1 score when regrading ConSE as project funtion, and related rules w.r.t unseen relations.

2974

4.5 Whether and how can logical rules help build better semantic space?

We investigate this question via logic rule analysis. General inspection of Table 2 reveals that rule-based embedding is slightly better than single KG embedding, and *KG+Rule embedding* achieves the best result with 3∼4% improvement in overall Hit@1 score under ConSE. A further examination of case studies over different kinds of semantic representations is listed in Table 4. It shows that most relations based on rule embedding achieve at least comparable results with KG embedding such as *nominated_for*, *producer*, *lyrics_by*. Some relations, such as *producer*, outperform KG embedding slightly. This may because rule embedding can capture logic-level connections between seen and unseen relations. For example, the unseen relation *nominated_for* is logically related with two seen relations *award_received* and *winner* with the rule *nominated_for(x,z)* ⟸ *award_received(x,y)* ∧ *winner(y,z)*. The most interesting aspect is about the relation *mother* for which KG embedding fails to compare with word embedding because of being poorly trained, while rule embedding achieves comparable scores with word embedding. The reason may be that rule embedding helps strengthen the embedding by incorporating more knowledge from related relations contained in the rules, thus making a correction to the relation embedding.
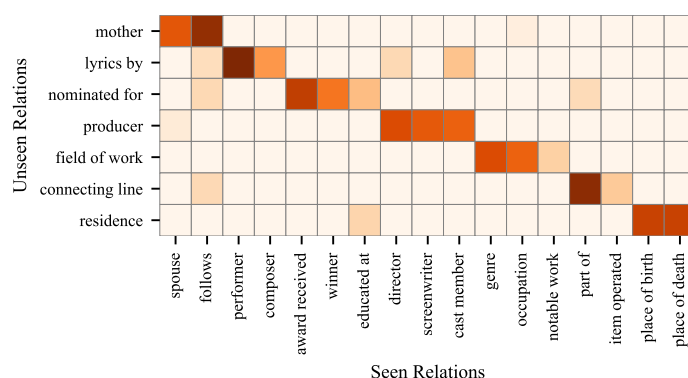

Figure 3: This heatmap is constructed from the result of ConSE+KG, reflecting the incidence of seen relations on unseen relations. Where the horizontal axis represents seen classes and the vertical axis represents unseen classes.

We also represent a heatmap of ConSE+KG results corresponding to Table 4, as shown in Figure3. From the heatmap, we can discover that KG embeddings could capture logical connections for the semantic connections between unseen and seen relations. For example, the relation *award_received* plays an important role in prediction of the unseen relation *nominated_for*, which is exactly consistent with the rule *award_received(x,y)* ∧ *winner(y,z)* ⟹ *nominated_for(x,z)* in Table 4. Similar matched correspondence for other relations are found. These results and analysis show that *logical connections between relations expressed by rules could help build right and explicit connections between unseen and seen relations, thereby building a better semantic space for ZSRC task.*

5 Conclusion and Future Work

We have studied the zero-shot relation classification task and took the first step towards bridging symbolic reasoning with semantic representations. Extensive experiments demonstrate the efficacy of our approach, revealing the advantages of knowledge graph embeddings and rules. In the future, we plan to exploit more efficient approaches to obtain symbolic rules an build end-to-end reasoning approaches for zero-shot tasks.

Acknowledgments

References

Z Akata, F Perronnin, Z Harchaoui, and C Schmid. 2016. Label-embedding for image classification. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 38(7):1425–1438.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*, pages 2787–2795.

Yang Chen, Daisy Zhe Wang, and Sean Goldberg. 2016. Scalekb: scalable learning and inference over large knowledge bases. *VLDB J.*, 25(6):893–918.

Shumin Deng, Ningyu Zhang, Jiaojian Kang, Yichi Zhang, Wei Zhang, and Huajun Chen. 2020a. Meta-learning with dynamic-memory-based prototypical network for few-shot event detection. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 151–159.

Shumin Deng, Ningyu Zhang, Zhanlin Sun, Jiaoyan Chen, and Huajun Chen. 2020b. When low resource nlp meets unsupervised language model: Meta-pretraining then meta-learning for few-shot text classification (student abstract). In *AAAI*, pages 13773–13774.

Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. 2018. Convolutional 2d knowledge graph embeddings. In *AAAI*, pages 1811–1818. AAAI Press.

A. Farhadi, I. Endres, D. Hoiem, and D. Forsyth. 2009. Describing objects by their attributes. In *IEEE Conference on Computer Vision & Pattern Recognition*.

Andrea Frome, Greg S Corrado, Jon Shlens, Samy Bengio, Jeff Dean, Tomas Mikolov, et al. 2013. Devise: A deep visual-semantic embedding model. In *Advances in neural information processing systems*, pages 2121–2129.

Luis Antonio Galárraga, Christina Teflioudi, Katja Hose, and Fabian Suchanek. 2013. Amie: association rule mining under incomplete evidence in ontological knowledge bases. In *Proceedings of the 22nd international conference on World Wide Web*, pages 413–422. ACM.

Luis Galárraga, Christina Teflioudi, Katja Hose, and Fabian M. Suchanek. 2015. Fast rule mining in ontological knowledge bases with AMIE+. *VLDB J.*, 24(6):707–730.

Zhou GuoDong, Su Jian, Zhang Jie, and Zhang Min. 2005. Exploring various knowledge in relation extraction. In *Proceedings of the 43rd annual meeting on association for computational linguistics*, pages 427–434. Association for Computational Linguistics.

Vinh Thinh Ho, Daria Stepanova, Mohamed H. Gad-Elrab, Evgeny Kharlamov, and Gerhard Weikum. 2018. Rule learning from knowledge graphs guided by embedding models. In *International Semantic Web Conference (1)*, volume 11136 of *Lecture Notes in Computer Science*, pages 72–90. Springer.

Lifu Huang, Heng Ji, Kyunghyun Cho, and Clare R Voss. 2017. Zero-shot transfer learning for event extraction. *arXiv preprint arXiv:1707.01066*.

Dinesh Jayaraman and Kristen Grauman. 2014. Zero shot recognition with unreliable attributes. In *International Conference on Neural Information Processing Systems*.

Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge graph embedding via dynamic mapping matrix. In *ACL (1)*, pages 687–696. The Association for Computer Linguistics.

Omer Levy, Minjoon Seo, Eunsol Choi, and Luke Zettlemoyer. 2017. Zero-shot relation extraction via reading comprehension. *arXiv preprint arXiv:1706.04115*.

Jingjing Li, Mengmeng Jing, Ke Lu, Lei Zhu, Yang Yang, and Zi Huang. 2019. From zero-shot learning to cold-start recommendation. *Age*, 25:30.

Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, and Song Liu. 2015a. Modeling relation paths for representation learning of knowledge bases. *arXiv preprint arXiv:1506.00379*.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015b. Learning entity and relation embeddings for knowledge graph completion. In *AAAI*, pages 2181–2187. AAAI Press.

Hanxiao Liu, Yuexin Wu, and Yiming Yang. 2017. Analogical inference for multi-relational embeddings. In *ICML*, volume 70 of *Proceedings of Machine Learning Research*, pages 2168–2178. PMLR.

Yi Luan, Luheng He, Mari Ostendorf, and Hannaneh Hajishirzi. 2018. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In *EMNLP*, pages 3219–3232. Association for Computational Linguistics.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In Christopher J. C. Burges, Léon Bottou, Zoubin Ghahramani, and Kilian Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 3111–3119.

Maximilian Nickel, Lorenzo Rosasco, and Tomaso A. Poggio. 2016. Holographic embeddings of knowledge graphs. In *AAAI*, pages 1955–1961. AAAI Press.

Mohammad Norouzi, Tomas Mikolov, Samy Bengio, Yoram Singer, Jonathon Shlens, Andrea Frome, Greg S Corrado, and Jeffrey Dean. 2013. Zero-shot learning by convex combination of semantic embeddings. *arXiv preprint arXiv:1312.5650.*

Abiola Obamuyide and Andreas Vlachos. 2018. Zero-shot relation classification as textual entailment. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pages 72–78.

Pouya Ghiasnezhad Omran, Kewen Wang, and Zhe Wang. 2018. Scalable rule learning via learning representation. In *IJCAI*, pages 2149–2155. ijcai.org.

Mark Palatucci, Dean Pomerleau, Geoffrey E. Hinton, and Tom M. Mitchell. 2009. Zero-shot learning with semantic output codes. In *International Conference on Neural Information Processing Systems*.

Pushpankar Kumar Pushp and Muktabh Mayank Srivastava. 2017. Train once, test anywhere: Zero-shot learning for text classification.

Pengda Qin, Xin Wang, Wenhu Chen, Chunyun Zhang, Weiran Xu, and William Yang Wang. 2020. Generative adversarial zero-shot relational learning for knowledge graphs. In *AAAI*, pages 8673–8680. AAAI Press.

Shruti Rijhwani, Jiateng Xie, Graham Neubig, and Jaime Carbonell. 2018. Zero-shot neural transfer for cross-lingual entity linking.

Bernardino Romera-Paredes and Philip Torr. 2015. An embarrassingly simple approach to zero-shot learning. In *International Conference on Machine Learning*, pages 2152–2161.

Ali Sadeghian, Mohammadreza Armandpour, Patrick Ding, and Daisy Zhe Wang. 2019. DRUM: end-to-end differentiable rule mining on knowledge graphs. In *NeurIPS*, pages 15321–15331.

Michael Sejr Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *ESWC*, volume 10843 of *Lecture Notes in Computer Science*, pages 593–607. Springer.

Daniil Sorokin and Iryna Gurevych. 2017. Context-aware representations for knowledge base relation extraction. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1784–1789.

Zhichun Wang and Juan-Zi Li. 2015. Rdf2rules: Learning rules from RDF knowledge bases by mining frequent predicate cycles. *CoRR*, abs/1512.07734.

Zifeng Wang, Rui Wen, Xi Chen, Shao-Lun Huang, Ningyu Zhang, and Yefeng Zheng. 2020. Finding influential instances for distantly supervised relation extraction. *arXiv preprint arXiv:2009.09841.*

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575.*

Hongbin Ye, Ningyu Zhang, Shumin Deng, Mosha Chen, Chuanqi Tan, Fei Huang, and Huajun Chen. 2020. Contrastive triple extraction with generative transformer. *arXiv preprint arXiv:2009.06207.*

Mo Yu, Wenpeng Yin, Kazi Saidul Hasan, Cícero Nogueira dos Santos, Bing Xiang, and Bowen Zhou. 2017. Improved neural relation detection for knowledge base question answering. In *ACL (1)*, pages 571–581. Association for Computational Linguistics.

Haiyang Yu, Ningyu Zhang, Shumin Deng, Zonggang Yuan, Yantao Jia, and Huajun Chen. 2020. The devil is the classifier: Investigating long tail relation classification with decoupling analysis. *arXiv preprint arXiv:2009.07022.*

Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao. 2015. Distant supervision for relation extraction via piece-wise convolutional neural networks. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1753–1762.

Ningyu Zhang, Shumin Deng, Zhanlin Sun, Xi Chen, Wei Zhang, and Huajun Chen. 2018. Attention-based capsule networks with dynamic routing for relation extraction. *arXiv preprint arXiv:1812.11321*.

Ningyu Zhang, Shumin Deng, Zhanlin Sun, Guanying Wang, Xi Chen, Wei Zhang, and Huajun Chen. 2019a. Long-tail relation extraction via knowledge graph embeddings and graph convolution networks. In *Proceedings of the NAACL*, pages 3016–3025.

Wen Zhang, Bibek Paudel, Liang Wang, Jiaoyan Chen, Hai Zhu, Wei Zhang, Abraham Bernstein, and Huajun Chen. 2019b. Iteratively learning embeddings and rules for knowledge graph reasoning. In *The World Wide Web Conference*, pages 2366–2377. ACM.

Ningyu Zhang, Shumin Deng, Zhanlin Sun, Jiaoyan Chen, Wei Zhang, and Huajun Chen. 2020a. Relation adversarial network for low resource knowledge graph completion. In *Proceedings of The Web Conference 2020*, pages 1–12.

Ningyu Zhang, Luoqiu Li, Shumin Deng, Haiyang Yu, Xu Cheng, Wei Zhang, and Huajun Chen. 2020b. Can fine-tuning pre-trained models lead to perfect nlp? a study of the generalizability of relation extraction. *arXiv preprint arXiv:2009.06206*.

Suncong Zheng, Yuexing Hao, Dongyuan Lu, Hongyun Bao, Jiaming Xu, Hongwei Hao, and Bo Xu. 2017. Joint entity and relation extraction based on a hybrid neural network. *Neurocomputing*, 257:59–66.

Ben Zhou, Daniel Khashabi, Chen-Tse Tsai, and Dan Roth. 2018. Zero-shot open entity typing as type-compatible grounding. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2065–2076.