FZR: Enhancing Knowledge Transfer via Shared Factors Composition in Zero-Shot Relational Learning

Zhijun Dong*
zjdong@mail.ustc.edu.cn
State Key Laboratory of Cognitive Intelligence, University of Science and Technology of China
Hefei, Anhui, China

Likang Wu*
wulk@tju.edu.cn
College of Management and Economics, Tianjin University
Tianjin, China

Kai Zhang†
kkzhang08@ustc.edu.cn
State Key Laboratory of Cognitive Intelligence, University of Science and Technology of China
Hefei, Anhui, China

Ye Liu
liuyer@mail.ustc.edu.cn
State Key Laboratory of Cognitive Intelligence, University of Science and Technology of China
Hefei, Anhui, China

Yanghai Zhang
yhzhang0612@mail.ustc.edu.cn
State Key Laboratory of Cognitive Intelligence, University of Science and Technology of China
Hefei, Anhui, China

Zhi Li
zhilizl@sz.tsinghua.edu.cn
Shenzhen International Graduate School, Tsinghua University
Shenzhen, Guangdong, China

Hongke Zhao
hongke@tju.edu.cn
Tanjin University
Tianjin, China

Enhong Chen
cheneh@ustc.edu.cn
State Key Laboratory of Cognitive Intelligence, University of Science and Technology of China
Hefei, Anhui, China

Abstract

Zero-Shot Relational Learning (ZSRL), strives to predict relations that have not been observed during training, presenting a considerable challenge in terms of model generalization. Existing ZSRL methods usually utilize the prior knowledge of labels (e.g., text description, ontological schema) to enable knowledge transfer by learned features. Nonetheless, these methods remain limited to calculating the surface features exhibited by relations, failing to fully explore their underlying driving factors. This leads to insufficient discrimination between the shared and distinctive inherent components among relations, which consequently impedes the cognitive understanding required for advanced reasoning. In our study, we aim to identify and utilize shared factors that widely exist in the prior knowledge of classes to learn enhanced semantic representations via shared factors composition, and develop our Factor-based ZSRL framework (FZR) with Generative Adversarial Networks (GANs) to bridge inequality between seen and unseen classes. FZR is designed to restructure the semantic space in such a way that it captures the essence of relation formation, thereby facilitating superior knowledge transfer in zero-shot scenarios. We conduct extensive experiments and evaluate our model on real-world datasets, and the results clearly demonstrate the effectiveness of the proposed model in zero-shot relational learning tasks.

CCS Concepts

• **Information systems → Data mining**; • **Computing methodologies → Knowledge representation and reasoning**; **Semantic networks**.

Keywords

Zero-Shot Relational Learning, Knowledge Transfer, Shared Factors, Generative Adversarial Networks

ACM Reference Format:
Zhijun Dong, Likang Wu, Kai Zhang, Ye Liu, Yanghai Zhang, Zhi Li, Hongke Zhao, and Enhong Chen. 2024. FZR: Enhancing Knowledge Transfer via Shared Factors Composition in Zero-Shot Relational Learning. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM '24), October 21–25, 2024, Boise, ID, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3627673.3679770

1 Introduction

Zero-Shot Learning (ZSL) represents a paradigm shift in machine learning, enabling models to classify and predict instances that fall outside their training scopes. Traditionally, ZSL has demonstrated versatility across tasks [39], including image classification [40], relation extraction [23], and node classification [37, 43]. However, the core challenge of knowledge transfer from familiar to novel classes remains. This issue is especially pronounced in the realm of knowledge graphs, where complex relational networks require a

*Equal Contributions.
†Corresponding Author.

CIKM '24, October 21–25, 2024, Boise, ID, USA

ACM ISBN 979-8-4007-0436-9/24/10
https://doi.org/10.1145/3627673.3679770

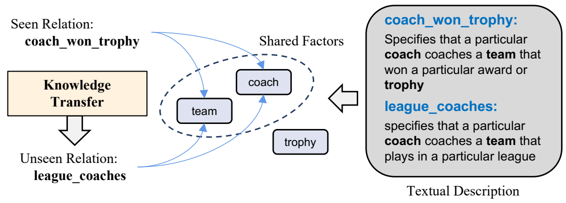


Figure 1: The knowledge transfer process of Zero-Shot Relational Learning with inherent shared factors.

depth of understanding that transcends mere superficial features. By representing semantic relationships between classes in a knowledge graph, such as hierarchical structures, associative properties, or other related information, this knowledge can be incorporated into the model for reasoning on unseen classes. The knowledge graph provides prior knowledge to the model, enabling better generalization in the inference process for new classes.

In the context of relational learning, ZSL offers a strategy to overcome the limitation of missing training samples by leveraging prior knowledge about relations. This technique relies on the exploitation of prior knowledge, such as textual descriptions [30] of relations or the ontological schemes [12] from Knowledge Graph (KG), to infer unseen relations. The underlying premise of this approach suggests that prior knowledge embodies the fundamental aspects of relations, thus establishing a platform for subsequent learning.

However, current methods in Zero-Shot Relational Learning (ZSRL) often focus on superficial aspects of relations and do not thoroughly investigate the underlying factors that drive them. This approach does not align with cognitive psychology's processes for recognizing new objects [34]. As a result, these methods do not adequately differentiate between shared and unique components of relations, hindering the cognitive understanding needed for more advanced reasoning. Additionally, there is a lack of research on optimally configuring sampling and learning strategies at the sample level within the ensemble framework [51]. As shown in Figure 1, It is manifest that the shared factors play a pivotal role in representing the associations between relations such as *'coach_won_trophy'* and *'leagues_coaches'*. These underlying elements are crucial for capturing the essence of the relational linkage.

The exploration of shared factors in ZSRL tasks is constrained by several limitations, preventing the establishment of well-defined paradigms for their effective use in representation learning. Firstly, there is a significant lack of methodological frameworks for systematically extracting and utilizing shared factors from relational text. Secondly, integrating these shared factors into the ZSRL framework poses a considerable challenge. Additionally, the presence of homogeneous relations within knowledge graphs (KGs) often results in highly similar shared factor representations, which diminishes the model's capacity for generalization. Therefore, improving the model's ability to distinguish between such homogeneous relations is a critical concern that requires attention.

To address these limitations, in this paper, we propose the Factor-based Zero-shot Relational Learning model (FZR), a novel framework designed to identify and utilize shared factors that widely exist in the prior knowledge of relations through factor extraction and shared factors composition. Then we employ a Generative Adversarial Network (GAN) to learn to embody these factors, enabling knowledge transfer between seen and unseen relations. In a manner similar to the situation encountered in [48, 49], although the approach differs, we introduce an expert-guided mechanism for reconstructing semantic representations of relations, where experts score the degree of relevance between relation pairs. This reconstruction is sensitive to the nuances among homogeneous relations and, when combined with the latent shared factors, results in an enhanced relation representation. Finally, to empirically validate the efficacy of our proposed model, we conduct extensive experiments on large-scale real-world datasets. Our experiments not only demonstrate the model's superior performance in predicting unseen relations but also highlight its ability to generalize from limited information, thereby setting a new precedent for future ZSL applications within knowledge graphs and beyond. In summary, the main contributions could be summarized as follows:

- We present the factor-based zero-shot relational learning framework (FZR) to learn enhanced semantic representations via shared factors composition.
- By integrating expert scores that assess the degree of relevance between homogeneous relations, we have reconstructed the semantic representation space. This ensures the model's sensitivity to subtle distinctions.
- We benchmark our model against state-of-the-art baselines on NELL-ZS and Wiki-ZS datasets. The comparative experiments clearly demonstrate the superiority of our model in zero-shot learning scenarios.

2 Related work

2.1 Zero-Shot Learning

The core of Zero-shot Learning (ZSL) is enabling knowledge sharing and inductive transfer between the seen and unseen classes. Traditional zero-shot learning paradigms predominantly leverage an attribute [22, 44, 45] or semantic space [9, 53] to establish interclass correlations. Within the attribute-based representation, some works endeavor to construct a correlation matrix between classes and attributes [1], while others utilize class attribute descriptions as a prior knowledge [8, 22, 24, 25], employing attributes to characterize class features. Such approaches learn semantic connections between classes through shared attributes, although they are notably sensitive to the quality of attributes and encounter domain shift challenges while attributes are tailored to describe localized class features. In the context of text-based semantic representation, some studies harness textual descriptions of classes [9, 17] or graph-like structures of classes [47] or pre-trained distributed word embeddings for modeling class semantic [27, 30, 50], deriving class semantic representations directly from widely available unstructured class text information [29]. To mitigate the pervasive noise in raw text, some employ TF-IDF features to diminish the influence of irrelevant words [41]. Furthermore, knowledge graph embedding-based methods [13, 31, 42] have been explored to learn semantic vectors within well-constructed KGs. However, the content of many large-scale public knowledge graphs, such as ConceptNet and WordNet, is not exhaustive, offering limited coverage of domain-specific knowledge. Besides, with the help of ontology-bsed knowledge representation and semantic information [12, 14], recent works explore richer and more competitive structured prior knowledge to model the inter-class relation for knowledge transfer in ZSL.

To achieve knowledge transfer from seen to unseen classes, a typical strategy is to learn a mapping function. Some works train class mapping based on instances of seen classes [6, 9, 22], and during testing, the input class vectors are projected into the corresponding vector space to identify the nearest neighbors as prediction labels [20]. Others learn an inverse mapping, projecting labels into the input instance space [5, 18, 42, 50], or propose aligning feature vectors and class instances within a unified space [10]. Nonetheless, trained on samples of seen classes, mapping-based approaches are more sensitive to seen classes and show weaker predictive performance for unseen classes. In recent years, generative models like GANs [15] have been proposed to generate samples for unseen classes based on class prior knowledge [24, 30, 44, 53, 54], transforming the ZSL task into a conventional supervised learning task and alleviating the issue of sparse samples for unseen classes. Although generative methods are also trained on samples of seen classes, the generator can enhance generalization performance for unseen classes based on semantic interlinkage between classes. In this paper, we also employ a GAN to construct our ZSL framework for inter-class knowledge transfer learning.

2.2 Zero-Shot Relational Learning

The majority of research works in Zero-Shot Learning (ZSL) have predominantly concentrated on various challenges within the realm of computer vision [11, 18, 42]. A subset of this research has expanded the application of ZSL to the task of knowledge graph completion, predicting the existence of hitherto unobserved entities by leveraging auxiliary linkages with previously recognized entities [16, 33, 38, 52], incorporating textual descriptions of entities [32] or learning entity-independent graph representations to naturally generalizing to unseen entities [7, 35]. Recently, the task of Zero-shot relational learning has attracted some attention, which involves introducing textual descriptions of relations [30] or ontological scheme [12, 14] to implement knowledge transfer between seen and unseen relations for predicting new-added facts. However, these methods remain limited to calculating the surface features of relations, failing to fully explore the underlying driving factors, creating a bottleneck in classification performance. In this paper, we propose a novel perspective to exploit the prior knowledge to obtain enhanced relation semantics.

3 Methodology

In this section, we first illustrate our research problem formally and present task-related notations. Afterwards we will introduce our proposed general ZSL framework FZR, which includes enhanced relation representation with shared factors composition, expert-guided semantic reconstruction and a generative adversarial network (GAN) to generate semantic embeddings for knowledge transfer learning. Figure 2 presents the overall framework of FZR.

3.1 Task Formulation

We focus on the specific zero-shot relational learning problem in KG link prediction task. The object of learning and fitting for the model is the embeddings of relation within the KG. We endeavor to construct enhanced relation representations, employing generative models to synthesize samples for unseen relations, thereby facilitating link prediction performance of these relations within general zero-shot relational learning setting.

In this task, a KG $\mathcal{G}$ is composed of a set of entities $\mathcal{E}$, a set of relations $\mathcal{R}$ and a set of triple facts $\mathcal{T} = \{(h, r, t)|h, t \in \mathcal{E}; r \in \mathcal{R}\}$. We use R_s, R_u to denote the set of seen relations and unseen relations, respectively. Note that, here $R_s \cap R_u = \emptyset$. Since our target can be formulated as predicting the tail entity e_2 given the head entity e_1 and the query relation r, for each query tuple (e_1, r), there are a ground-truth tail entity e_2 and a candidate set $C(e_1, r)$. Then we define $D_{tr} = \{(e_1, r_s, e_2, C(e_1, r_s))|e_1, e_2 \in \mathcal{E}; r_s \in R_s\}$ as the training dataset. During testing, the proposed model is to predict the relational facts of unseen relations $r_u \in R_u$. It is noted that we consider a closed set of entities. Specifically, each entity that appears in the testing triples is still in the entity set $\mathcal{E}$. Thus, we define $D_{te} = \{(e_1, r_u, e_2, C(e_1, r_u))|e_1, e_2 \in \mathcal{E}; r_u \in R_u\}$ as the testing dataset. In this setting, zero-shot link prediction aims to assign the highest ranking score to e_2 against the rest of all the candidate entities in $C(e_1, r_u)$. Therefore, during testing, we will predict the triple facts of r_u by ranking e_2 with the candidate tail entities $e_2^{'} \in C(e_1, r_u)$.

3.2 Enhanced Relation Representation with Shared Factors Composition

In cognitive psychology [34], it has been observed that multiple small units collaboratively engage in a series of cognitive activities. When humans encounter new procedural knowledge, they update their cognitive activities by updating related components across these units [21]. In our study, we propose that shared factors serve as crucial units in facilitating new relation learning, and introduce a novel methodology termed Shared Factors Composition (SFC). The SFC is designed to construct enhanced semantic representations of shared factors from the textual descriptions of relations, thereby amplifying the effectiveness of knowledge transfer within zero-shot relational learning settings. Considering this, our SFC approach innovatively combines keyphrase extraction, clustering algorithms, and the composition of factors representations, which provides a comprehensive framework for representing relations in zero-shot relational learning, augmenting the model's comprehension of unseen relations, and offering an effective avenue for exploring the latent interconnections among relations.

With the text descriptions as prior knowledge, the initial semantic representations of relations can be obtained by the pre-trained word embedding model, Glove [28]. For a given text description, composed of words $\{w_1, w_2, \ldots, w_n\}$, the initial semantic representation $\mathbf{r}_o$ of relation r is the aggregate of its word vectors:

$$\mathbf{r}_o = \frac{1}{n}\sum_{i=1}^{n}\mathbf{w}_i \tag{1}$$

To lay the groundwork for SFC, we initiate the process with the extraction of key factors from text descriptions, which serve as the

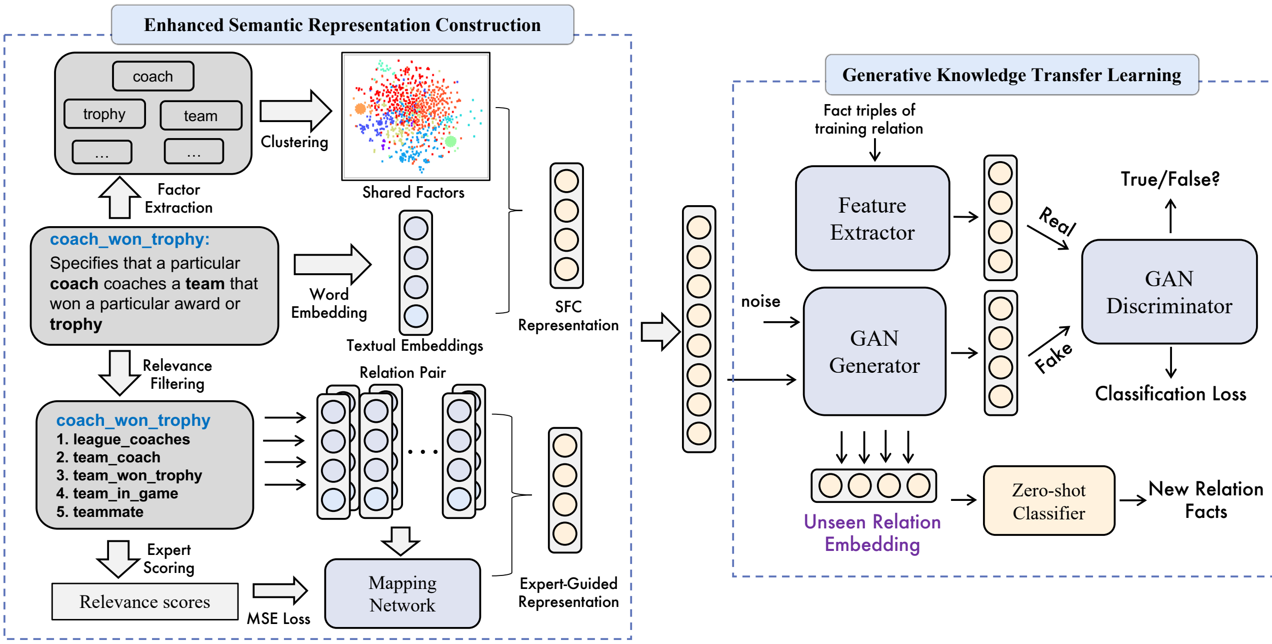


Figure 2: Overview of the proposed ZSL framework FZR. In detail, the two panels indicate the enhanced semantic representation construction process and generative knowledge transfer learning with GAN, respectively.

foundational elements of our factor library. For this pivotal part, we leverage the unsupervised keyword extraction algorithm provided by Python Keyphrase Extraction toolkit [3] to efficiently identify the most salient phrases and terms that capture the essence of the textual content, which are then aggregated to form a shared factor library. It is imperative to note that, the extracted factors are not subjected to deduplication, thereby preserving the integrity of their individual contributions to the overall semantic structure.

Within the context of prior knowledge, the semantic interconnectivity inherent in relations is discernible. Even though the deployment of these relations within disparate domains, the semantic interlinkage is capable of being apprehended and integrated by models to bolster inferential reasoning. Hence, the method we introduce is dedicated to uncovering such shared semantic factors among relations, and to the reconstruction of the relations' initial semantic representations $\mathbf{r}_o$ through shared factors composition. Specifically, SFC is accomplished by clustering within the representation of factor library, thereby ensuring that the emergent cluster representations encapsulate the semantic properties of the shared factors that widely exist in KG relations. In this endeavor, we also proposed to employ unsupervised clustering algorithms(e.g.,K-means [26]) to affirm the robustness of our method. By calculating the similarity between each initial semantic representation of relation and the representations of cluster centroids $\{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_k\}$, we can assign a set of weights to each relation, indicating the importance of each shared factor in the new semantic representation:

$$\alpha_i = \frac{\exp(\cos(\mathbf{r}_o, \mathbf{c}_i))}{\sum_{i=1}^{k} \exp(\cos(\mathbf{r}_o, \mathbf{c}_i))} \tag{2}$$

where *cos* denotes the cosine similarity. And then we get the reconstructed representation $\mathbf{r}_c$ as the SFC representation:

$$\mathbf{r}_c = \sum_{i=1}^{k} \alpha_i \mathbf{c}_i \tag{3}$$

3.3 Expert-Guided Semantic Reconstruction

Our innovation recognizes that domain-specific knowledge graphs (KGs) often contain many homogeneous relations, which, while similar in factor representations, require careful differentiation. For example, the relations purchase' and invest' both involve the concept of 'buying something,' but they apply to different domains and involve different types of entities. To address this, we introduce an expert-guided semantic reconstruction mechanism that redefines the relational semantic space with precision. By incorporating expert knowledge through manually assigned relevance scores between similar relations, we enhance the model's ability to distinguish between them during reasoning.

3.3.1 Preliminary Filtering Mechanism. It is essential to recognize the profusion of relations that populate real-world knowledge graphs, a factor that renders the exhaustive manual scoring of each relation pair an unfeasible task. To address this challenge, a more pragmatic approach is suggested. Domain experts are invited to direct their evaluative expertise selectively towards those relation pairs that have been pre-identified as relevant through the computational determination of similarity in relation embeddings. This preliminary filtering mechanism is designed to enhance the efficiency of the expert scoring process, allowing for a concentrated focus on relation pairs that are deemed significant based on their computed semantic proximity. This targeted approach not only streamlines the evaluation process but also ensures that the semantic reconstruction is grounded in the most pertinent and semantically rich relational data, as determined by expert judgment.

Let $\mathbf{R}$ be the set of all relation pairs in the knowledge graph, and let $\mathbf{r}_i$ and $\mathbf{r}_j$ be the embeddings of two relations i and j respectively. We define the similarity function $S : \mathbf{R} \times \mathbf{R} \to \mathbb{R}$, which computes the similarity between any two relation embeddings. The function can be, for instance, the cosine similarity:

$$S(\mathbf{r}_i, \mathbf{r}_j) = \frac{\mathbf{r}_i \cdot \mathbf{r}_j}{\|\mathbf{r}_i\| \|\mathbf{r}_j\|} \tag{4}$$

Given a threshold θ, the preliminary filtering mechanism $\mathcal{F}$ selects a subset of relation pairs $\mathbf{R}' \subseteq \mathbf{R}$ deemed relevant:

$$\mathbf{R}' = \mathcal{F}(\mathbf{R}, \theta) = \{(\mathbf{r}_i, \mathbf{r}_j) \in \mathbf{R} \mid S(\mathbf{r}_i, \mathbf{r}_j) \geq \theta\} \tag{5}$$

For a given relation $\mathbf{r}_i$, we define the recall function $C : \mathcal{R} \to 2^{\mathcal{R}}$, which maps $\mathbf{r}_i$ to its recall list of relations from $\mathbf{R}'$. For convenience, we set n to restrict the number of recalled relations.

$$C(\mathbf{r}_i) = \{\mathbf{r}_j \mid (\mathbf{r}_i, \mathbf{r}_j) \in \mathbf{R}' \land |C(\mathbf{r}_i)| \leq n\} \tag{6}$$

3.3.2 Semantic Space Reconstruction. In this system, domain experts provide scores for the relatedness between pairs of relationships, which are then incorporated into the reconstruction process as a key reference for reconstructing the semantic space of relations.

Let $E(\mathbf{r}_i, \mathbf{r}_j)$ be the expert score for the relevance between the representation $\mathbf{r}_i$ and another representation $\mathbf{r}_j$ in recall list $C(\mathbf{r}_i)$. We aim to learn a mapping function f such that the transformed representations reflect the expert-level semantic comprehension:

$$f(\mathbf{r}_i) = \arg\min_f \sum_{\mathbf{r}_j \in C(\mathbf{r}_i)} \|f(\mathbf{r}_i) - f(\mathbf{r}_j)\|_2^2 \cdot E(\mathbf{r}_i, \mathbf{r}_j) \tag{7}$$

Consequently, for each relation r , we are capable of learning two distinct types of representations, i.e., the shared factors composition representation as mentioned in 3.2 and the expert-guided semantic representation as mentioned in 3.3. To fuse the semantic features from these two representations, we concatenate them to form the final representation of the relation:

$$\mathbf{r}^* = [\mathbf{r}_c; f(\mathbf{r}_o)] \tag{8}$$

3.4 Generative Knowledge Transfer Learning

With the enhanced semantic representations, we next show how to utilize them for generative knowledge transfer learning. Specifically, we will introduce how we obtain the real embedding of relations by feature extractor and the framework of our generative method. Given the demonstrated effectiveness of Generative Adversarial Networks (GANs) in learning the congruence between class vectors and their corresponding instances, we employ GANs to facilitate the transfer of knowledge from seen to unseen relations. Our GAN architecture comprises three integral components: a generator G, a feature extractor E, and a discriminator D.

3.4.1 Feature Extractor. In contrast to traditional knowledge graph embedding methods, which learn entity and relation embeddings based on certain assumptions or constraints, our approach aims to learn the cluster-structured feature distribution of both seen and unseen relational facts, thereby preserving high intra-class similarity and relatively low inter-class similarity, in line with the majority of ZSL works. Building upon previous works [12, 30] for learning and training authentic relation embeddings, we posit the existence of an entity pair set $T_r = \{(e_1, e_2) | (e_1, r, e_2) \in \mathcal{T}\}$ where $r \in R_s$. An entity pair set T_r is denoted by the seen relation r and contains all triples involving the relation r. Consequently, the true embedding x_r of relation r is represented by the embedding of entity pairs within set T_r, thus allowing for supervised training via a selection of reference triples from this set.

Specifically, for the entity pair (e_1, e_2) in the set T_r, we first embed each entity using a simple Fully Connected (FC) layer, generating the embedding u_{ep} for the entity pair:

$$\begin{aligned} u_{ep} &= \sigma([f_1(v_{e_1}); f_1(v_{e_2})]) \\ f_1(v_e) &= W_1(v_e) + b_1 \end{aligned} \tag{9}$$

where $[.;.]$ denotes the concatenation operation, σ is *tanh* activation function. We further consider the one-hop structure of each entity, embedding the one-hop neighbors through an FC layer as well. For the head entity e_1, its structural embedding u_{e_1} is represented by the aggregation of embeddings from its neighbor nodes. The structural embedding u_{e_2} for the tail entity e_2 is computed in the same way.

$$\begin{aligned} u_e &= \sigma\Big(\frac{1}{\|\mathcal{N}_e\|} \sum_{(r^n, e^n) \in \mathcal{N}_e} f_2(v_{r^n}, v_{e^n})\Big) \\ f_2(v_{r^n}, v_{e^n}) &= W_2(v_{r^n}; v_{e^n}) + b_2 \end{aligned} \tag{10}$$

where $\mathcal{N}_e = (r^n, e^n) | (e, r^n, e^n) \in \mathcal{T}$ denotes the one-hop neighbors of entity e, f_2 is also a FC layer to encode the corresponding neighbor entity. The final embedding for the entity pair (i.e., the true embedding x_r of relation r) is represented by the concatenation of u_{ep}, u_{e_1} and u_{e_2}.

$$x_r = x_{(e_1, e_2)} = [u_{e_1}; u_{ep}; u_{e_2}] \tag{11}$$

We employ some reference triples to train the true relation embeddings. With each relation r, the entity pairs in set T_r are randomly divided into two parts: a reference set (e_1^*, r, e_2^*) and a positive set (e_1^+, r, e_2^+). We also generate a set of negative triples (e_1^+, r, e_2^-) by replacing the tail entity of each triple in the positive set with other entities. For m reference triples, we take the mean of the reference relation embeddings $x_{(e_1^*, e_2^*)}$ and calculate the cosine similarity with each positive triple's relation embedding and each negative triple's relation embedding as $score_\omega^+$ and $score_\omega^-$. A margin ranking loss is then employed to optimize the training:

$$L_\omega = max(0, \gamma + score_\omega^+ - score_\omega^-) \tag{12}$$

where $\omega = W_1, W_2, b_1, b_2$ is the parameter set and γ denotes the margin parameter.

3.4.2 Adversarial Training. During adversarial training, the enhanced relation embeddings serve as inputs to the generator G, concatenated with random noises. Real embeddings of relations can be provided by feature extractor E, and the discriminator D will distinguish the generated embeddings from the real ones. We generate sample embeddings instead of raw samples to ensure both accuracy and efficiency as in many works [12, 30, 44].

Formally, for a relation r_i, the generator G take as input its enhanced representation r_i^* and a random noise z sampled from

Table 1: Statistics of the ZSRL datasets. # Ent. and # Triples denote the number of entities and triples in KGs. # Train/Dev/Test denotes the number of relations in training/validation/testing sets.

Dataset	# Ent.	# Triples	# Train/Dev/Test
NELL-ZS	65,567	188,392	141/10/30
Wiki-ZS	605,812	724,967	469/20/48

a normal distribution, then generate its embeddings: $\hat{x} = G(z, r_i^*)$. Following the settings in [30], the loss of G is defined as:

$$L_G = -\mathbb{E}[D(\hat{x})] + \lambda_1 L_{cls}(\hat{x}) + \lambda_2 L_P \tag{13}$$

where the first term is Wasserstein loss to ensure the generated embeddings distribution closely aligns with the real embeddings distribution, the second term is a supervised classification loss for classifying the synthesized embeddings, and the third is for regularizing the mean of generated embeddings of each class to be the mean of its real embeddings. Both of the latter two loss terms encourage the generated embeddings to have more inter-class discriminability. λ_1 and λ_2 are the corresponding weight coefficients.

The discriminator D is tasked with distinguishing between synthetic embeddings $\hat{x}$ from G and real features x provided by the E. The loss function for the discriminator D consists of five components: the first two approximate the Wasserstein distance between the distributions of synthetic and real embeddings, the third and fourth are supervised classification loss, the fifth is a gradient penalty that ensures the gradients of D are of unit norm, which aids in stabilizing the training process and preventing mode collapse:

$$L_D = \mathbb{E}[D(x, r_i^*)] - \mathbb{E}[D(\hat{x})] + \lambda_3 L_{cls}(\hat{x}) + \lambda_4 L_{cls}(x) + L_{GP} \tag{14}$$

where λ_3 and λ_4 are the weight coefficients of the two supervised classification loss.

3.5 Implement Prediction for Unseen Relations

Leveraging the capabilities of a well-trained GAN, we utilize the generator G to synthesize features and implement task-specific prediction for classes that have not been encountered during training. The evaluation of triples is conducted by measuring the similarity between the generated embedding of relation r and the joint embedding of the entity pair (e_1, e_2).

Specifically, given a unseen relation r_u and its enhanced semantic representation r_u^*, the generator can synthesis plausible embedding $\hat{x}_{r_u} = G(z, r_u^*)$. For a query tuple (e_1, r_u), the similarity ranking score can be calculated by the cosine similarity between $\hat{x}_{r_u}$ and $x_{(e_1,e_2)}$ while $e_2 \in C(e_1, r_u)$:

$$score_{(e_1,r_u,e_2)} = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} score^i_{(e_1,r_u,e_2)} \tag{15}$$

where N_{test} denotes the number of generated embeddings of specific relation since we could set an arbitary number to satisfy demands of different tasks.

4 Experiments

4.1 Experimental Setup

4.1.1 Datasets. We conduct the experiments of our model and baselines on two public benchmarks proposed by [30], i.e., NELL-ZS and Wiki-ZS formulated from NELL [4] and Wikidata, respectively. Each relation in these datasets is accompanied by a textual description and the types of its corresponding entities. Within each knowledge graph, seen relations are included in the training set and validation set, while unseen relations are confined to the test set, with corresponding triples distributed accordingly.

Given to the lack of a detailed explanation for the dataset splits by the previous work [12, 30], and in the interest of adhering to a more general Zero-Shot Relational Learning setting as well as facilitating a fairer comparison, we implement a three-fold random partitioning of the Train/Dev/Test splits. Subsequently, we conduct three-fold cross-validation over these partitions. It is imperative to underscore that unlike the conservative approach of [12], which fixed the test relations and merely reallocated the training and validation sets, we posit that the model performance on the re-partitioned datasets more vividly reflects their adaptability to zero-shot scenarios and offers a more robust validation of their generalization capabilities. Therefore, we re-conducted the experiments for all baseline models on the re-partitioned datasets. More statistics about these datasets are presented in Table 1.

4.1.2 Evaluation Metrics. During testing, the models are employed to rank all tail entities from a candidate list given a head entity and a relation in a test triple. Consequently, following the setting in [30], we adopt commonly used metrics in KGC tasks: mean reciprocal ranking (MRR) and hits at ranks 10 (H@10), 5 (H@5), and 1 (H@1) for evaluation purposes. MRR represents the average of the reciprocal ranks of all correct entities, and H@k denotes the percentage of testing samples where the correct entity is ranked within the top k positions among all candidates.

4.1.3 Baselines. We compare our ZSL method with the state-of-the-art baselines, including three traditional Knowledge Graph Embedding (KGE) methods: TransE [2], DistMult [46], and ComplEx [36], along with three generative approaches for Zero-Shot Learning: ZSGAN [30], OntoZSL [12], and DOZSL [14].

Although traditional KGE methods are not specifically designed for zero-shot learning tasks, including them in comparative experiments is essential for establishing a baseline performance for comparison and analysis. Inspired by ZSGAN [30], we propose zero-shot variants of three KGE methods: ZS-TransE, ZS-DistMult, and ZS-ComplEx. Instead of using randomly initialized vectors to represent relations in triples, we employ a feed-forward network similar to our generator's architecture to obtain relation representations. The input to this network is the textual embedding of the relation, which is fine-tuned along with the entity embeddings during training. During testing, the textual embeddings of unseen relations are processed through this network to generate relation representations, and predictions for factual triples are made using the original methods' scoring functions.

Among the generative ZSL baselines, ZSGAN [30] acquires relation embedding through the textual description of relation, OntoZSL [12] obtains relation embedding by jointly encoding the

Table 2: Results(%) of zero-shot link prediction with unseen relations. The Bold results are the best performance.

Method type	Model	NELL-ZS				Wiki-ZS			
		MRR	Hit@1	Hit@5	Hit@10	MRR	Hit@1	Hit@5	Hit@10
Traditional KGE-based method	ZS-TransE	20.3	13.4	27.1	32.0	15.1	8.7	20.8	27.5
	ZS-DistMult	21.9	17.7	25.9	29.0	16.4	13.4	19.2	21.4
	ZS-ComplEx	21.6	17.1	26.0	29.4	11.4	7.5	15.1	18.2
Generative-based method with pre-trained TransE	ZSGAN	38.0	29.2	47.3	53.6	40.8	35.3	46.7	50.9
	OntoZSL	38.9	29.5	49.0	56.6	41.3	35.6	46.6	52.0
	DOZSL	39.7	30.0	50.4	57.7	41.6	35.6	47.3	52.4
	FZR(ours)	41.9	32.6	**52.1**	59.1	42.7	36.7	48.3	53.5
Generative-based method with pre-trained DistMult	ZSGAN	40.0	31.9	48.5	55.5	43.6	38.6	49.1	52.4
	OntoZSL	40.2	31.2	50.0	57.6	44.7	38.8	49.2	53.7
	DOZSL	41.1	32.3	51.2	58.3	44.6	**39.7**	48.9	53.3
	FZR(ours)	**43.3**	**34.9**	51.9	**59.5**	**45.1**	39.2	**50.0**	**54.6**

textual description and the ontological structural knowledge. Both of them subsequently employ a GAN to generate the corresponding relation representation. DOZSL [14], built upon OntoZSL, encodes the disentangled ontological structure of relation to obtain relation representation and introduces a propagation-based ZSL learner. In all generative ZSL approaches, a feature extractor is used in conjunction with pre-trained KG embeddings. For the sake of generalizability, we adopt two representative KGE models, TransE [2] and DistMult [46] within the feature extractor in our experiments.

4.1.4 Implementation Details. For the NELL-ZS dataset, we employ a feature extractor pre-trained with 100-dimensional KGE embeddings to obtain 200-dimensional relation embeddings. For the Wiki-ZS dataset, in accordance with the settings of [30], we utilize a feature extractor pre-trained with 50-dimensional KGE embeddings to acquire 100-dimensional relation embeddings.

The training of the feature encoder and generative model is conducted using the Adam optimizer [19] for parameter updates, with the margin parameter γ set to 10. For the feature encoder, the upper limit of neighbor count is set at 50, with a number of triples k considered for a training step fixed at 30, and the learning rate established at $\mathbf{5e^{-4}}$. For the generative model, the learning rate is set to $\mathbf{1e^{-3}}$, with β_1 and β_2 parameters set to 0.5 and 0.9, respectively. The discriminator is updated five times for every update of the generator. The dimension of the random vector $\mathbf{z}$ is 15, with the number of relation embeddings generated per batch set at 20. The weight for the classification loss λ_1, λ_3, λ_4 are set at 1, 0.5, 0.5. The pivot regularization λ_2 is set at 3.

For NELL-ZS, the generator's hidden units is set to 250, outputting 200-dimensional relation embeddings, while the discriminator's hidden units is set at 200, outputting a 2-dimensional vector to determine whether the input is real data. While for Wiki-ZS, the hidden units of the generator is set to 250 and that of the discriminator is 100. For relation text word embeddings, we utilize the published 300-dimensional word embeddings from glove.6B.300d.txt and employ TF-IDF to obtain denoised text embeddings. For each relation text description, we extract three key words as factors. For the corpus composed of all shared factors, we set the number of factor clusters to 10. For each relation, we recall five relations from all relations for relevance scoring.

4.2 Main Results

We report the results of ZSL testing on NELL-ZS and Wiki-ZS datasets in Table 2. From the results, we find that our FZR method (with TransE and DistMult) achieves a obvious lead compared with either traditional KGE-based methods or generative-based methods. This demonstrates the effectiveness of our factor-based zero-shot relation learning designs. Moreover, we can observe more interesting findings from this table:

First, there is a significant performance improvement for all generative-based methods over traditional KGE-based methods. To mitigate the performance fluctuations caused by different pre-trained KGE embeddings, we test each generative-based method with pre-trained TransE and DistMult embeddings, and observe significant performance improvements, underscoring the knowledge transfer capabilities of generative models. Second, among four generative-based methods, our FZR model significantly outperforms other methods with either pretrained TransE or DistMult. This proves that shared factors provide a more extensive linkage between seen and unseen relations, thus enabling more effective predictions for unseen relations. In more detail, compared to DOZSL on the NELL-ZS dataset using the same TransE pre-trained embeddings, FZR exhibited an increase of 2.6% in Hit@1 and 2.3% in Hit@5. Similarly, with DistMult pre-trained embeddings, our FZR model showed improvements of 2.6% in Hit@1 and 0.7% in Hit@5. Third, the performance improvements of our model on the Wiki-ZS dataset are less significant than NELL-ZS. This is likely due to the dataset's short relation descriptions, which provide limited prior knowledge, making the extracted factors less effective at explaining the relations. Consequently, the performance gap between our method and those that directly leverage the prior knowledge of relations, like ZSGAN, is not as pronounced.

From the perspective of performance improvement through model optimization, the method we used is very simple and efficient. As shown in Table 2, our method achieves significant improvements in model performance. For example, on the NELL-ZS dataset, FZR outperforms DOZSL by a larger margin (5.4%) in terms

Table 3: Results(%) of FZR on NELL-ZS when embeddings of shared factors composition("-sfc") and embeddings of expert-guided semantic("-exp") are removed from the enhanced representations of relations.

Test set	Model	MRR	Hit@1	Hit@5	Hit@10
NELL-ZS	FZR	**41.9**	**32.6**	**52.1**	**59.1**
	FZR(-sfc)	41.2	32.3	51.0	57.4
	FZR(-exp)	41.6	32.4	51.8	58.1
	FZR(-sfc&exp)	38.0	29.2	47.3	53.6

Table 4: Hit@1(%) of our hyperparameter tune-up experiments for the number of clusters for shared factors and the recalled relations (RR) scoring by experts.

Clusters \ RR	n=1	n=2	n=3	n=4	n=5	n=6	n=7
k=5	28.1	30.3	29.8	31.0	31.3	31.1	31.0
k=10	29.4	30.5	30.0	31.1	**32.6**	32.5	32.3
k=15	29.2	30.9	30.7	30.5	31.5	31.3	31.0
k=20	28.2	29.6	31.0	31.2	31.7	31.4	31.1

of MRR compared to the performance improvement of DOZSL over ZSGAN (2.8%). Additionally, the performance improvement of FZR over DOZSL in terms of Hit@1 (8.0%) is also higher than the performance improvement of DOZSL over OntoZSL (3.5%).

4.3 Ablation Studies

To further verify the effectiveness of the individual components constituting our enhanced relation representation, we seriously design ablation experiments. Specifically, the experiments encompass FZR (our full proposed model), FZR-sfc (FZR without shared factors composition representation), FZR-exp (FZR without the expert-guided semantic representation), and FZR-sfc&exp (FZR lacking both shared factors composition representation and expert-guided semantic representation, effectively equivalent to ZSGAN).

We train these models on the NELL-ZS dataset, using TransE as the pre-trained KG Embedding for training the feature extractor. The test results for each ablated variant are reported in Table 3. The results affirm the significance of both the expert-guided semantic representation and the shared factor composition in our FZR model. The complete FZR model, which incorporates these two components, consistently outperformed its ablated variants. The FZR(-exp) variant showed a slight decrease in performance, suggesting that while this component enhances the model's capabilities of generalization, it is not the sole contributor of FZR. Meanwhile, the FZR(-sfc) model, without the shared factor composition, faced a more substantial reduction in performance, highlighting its importance in the model's architecture. The combined absence of both components in the FZR(-sfc&exp) variant led to the most significant performance drop, confirming that the synergy between the expert-guided semantic representation and the shared factor composition is crucial for the model's ability to effectively perform zero-shot learning.

4.4 Hyperparameter Tune-up

In our model, two principal hyperparameters are of paramount importance: the number of clusters k for shared factors as mentioned in subsection 3.2 and the number of recall relations n employed for expert scoring as mentioned in subsection 3.3. These hyperparameters influence the performance of the shared factors composition representation and the expert-guided semantic representation modules within our FZR framework, respectively. This subsection is dedicated to an exploration of the impact that the settings of these hyperparameters on the performance of Zero-shot Relational Learning (ZSRL).

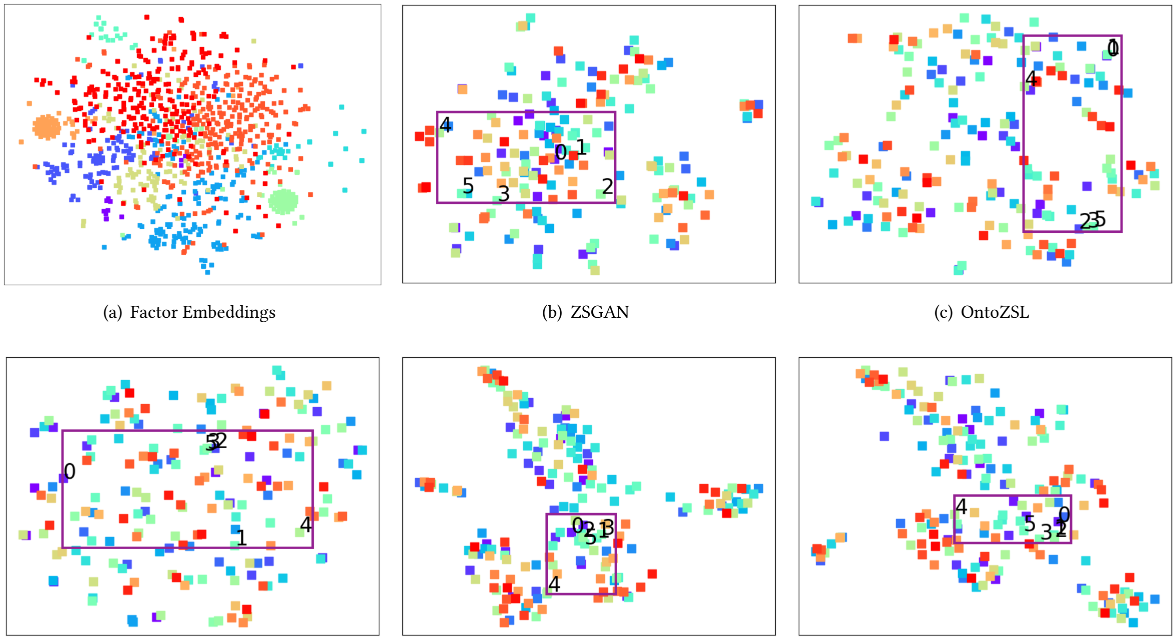
As illustrated in Table 4, we report the hit@1 performance for the FZR model on the NELL-ZS dataset, with the range of k set between [5, 10, 15, 20] and n between [1, 2, 3, 4, 5, 6, 7]. It is evident that the model achieves optimal performance when k is set to 10 and n to 5. Additionally, we observe an overall improving trend in performance as n increases when n is less than 5, which suggests that our scoring mechanism for recalled similar relations contributes positively to the improvement of the representational distribution of relations. We also observe that as n continues to increase, the model performance does not further improve[1]. We believe the choice of the value of n needs to consider the scale of the dataset and the distribution of homogeneous relations. For example, in the case in NELL-ZS datasets which includes 181 relations, there may not be a large number of homogeneous relations. Thus, even if we increase the value of n, it may not bring substantial performance improvements to the model. Concurrently, a smaller value of k may lead to a greater overlap within the semantic space, whereas an excessively large value of k can introduce superfluous information, which is detrimental to the learning of relations. This delicate balance between k and n highlights the critical need for meticulous tuning of these hyperparameters to enhance the ZSRL model's performance effectively.

4.5 Visualization of Relation Representation

4.5.1 Factor Embeddings Visualization. To provide a more intuitive and succinct presentation of qualitative results, we have visualized the clustering of factor representations derived from textual descriptions of relationships within the NELL-ZS dataset, as illustrated in Figure 3 (a). Here we set the number of cluster at 10, where each cluster is represented by a unique color. We can find that certain clusters exhibit a high density of closely associated factors, distinctly segregated from other clusters, while some clusters display a more scattered arrangement of factors. This suggests that our resultant shared factor representations encapsulate both the explicit and subtle features of relation semantics.

4.5.2 FZR Embeddings Visualization. As shown in Figure 3 (b) - (f), we present a visualization that contrasts the proposed representation of our method (r^* in Eq. (8)) against that of baseline

[1]We do not increase n further due to the expensive cost of expert labeling.



(a) Factor Embeddings (b) ZSGAN (c) OntoZSL

(d) DOZSL (e) FZR(sfc) (f) FZR

Figure 3: Visualization of factor embeddings and relation representation of baseline models and our proposed model FZR.

approaches. Specifically, Figure 3 (b) - (d) represent the representation of baselines; (e) indicates the representations of Shared Factors Composition (r_c in Eq. (3)); (f) represents our proposed representation in FZR model (r^* in Eq. (8)). Each color square in Figure 3 represents a relation. Due to the large number of relations in the dataset, we did not provide explanations for each color square. The purpose of these images is to demonstrate the distribution of relation representations learned by different models.

This visual comparison is designed to underscore the distinction in how the relation embeddings from our model capture inter-relation semantics when juxtaposed with those obtained from baseline methods. More specifically, we have showcased the embeddings of the *'language_of_university'* relation in conjunction with five other semantically proximate relations: *'academic_program_at_university'*, *'person_graduated_from_university'*, *'person_graduated_school'*, *'country_language'*, *'person_attends_school'*. These relations are annotated within the visualization as indices 0 through 5, respectively. The graphical representation clearly indicates that our model (FZR)'s embeddings coalesce more closely, reflecting a tighter semantic clustering of the six relations, as opposed to the more dispersed embeddings produced by the baseline models.

5 Conclusion

In this paper, we introduce a novel model for Zero-shot Relational Learning (ZSRL) named FZR, predicated on leveraging shared factors and expert knowledge to augment the transfer of more profound knowledge between relations. More specifically, we commence by discovering shared factors that widely exist in the prior knowledge of classes, learning enhanced relation representation by shared factors composition. Concurrently, we engage in expert scoring for homogeneous relations recalled through similarity calculation, and utilize the relevance scores between relations to reconstruct the semantic representation space, thereby yielding class-sensitive relational semantic representations. Employing these representations, we utilize a Generative Adversarial Network (GAN) to facilitate Knowledge Transfer Learning between seen and unseen relations. Upon training, our model demonstrates a heightened performance in predicting unseen relational facts. Extensive experiments of our model conducted on real-world datasets have yielded performance metrics that surpass several state-of-the-art methods. Hence, the acquisition of shared factors and the construction of class-sensitive semantics could emerge as effective ways to enhance relation representations. The further consideration of integrating structured information between relations, to facilitate real-time model updates through incremental learning in the face of limited prior knowledge, stands as a promising direction.

6 Acknowledgments

This research was partial supported by grants from the National Key Research, Development Program of China (Grant No. 2021YFF090 1000), the National Natural Science Foundation of China (Grants No. U20A20229, 72101176, 62206155), Anhui Provincial Natural Science Foundation (No. 2308085QF229), and the Fundamental Research Funds for the Central Universities (No. WK2150110034).

References

[1] Zeynep Akata, Florent Perronnin, Zaid Harchaoui, and Cordelia Schmid. 2013. Label-embedding for attribute-based classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 819–826.

[2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems* 26 (2013).

[3] Florian Boudin. 2016. pke: an open source python-based keyphrase extraction toolkit. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*. Osaka, Japan, 69–73. http://aclweb.org/anthology/C16-2015

[4] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam Hruschka, and Tom Mitchell. 2010. Toward an architecture for never-ending language learning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 24. 1306–1313.

[5] Soravit Changpinyo, Wei-Lun Chao, Boqing Gong, and Fei Sha. 2016. Synthesized classifiers for zero-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5327–5336.

[6] Jiaoyan Chen, Freddy Lécué, Yuxia Geng, Jeff Z Pan, and Huajun Chen. 2020. Ontology-guided semantic composition for zero-shot learning. *arXiv preprint arXiv:2006.16917* (2020).

[7] Mingyang Chen, Wen Zhang, Yushan Zhu, Hongting Zhou, Zonggang Yuan, Changliang Xu, and Huajun Chen. 2022. Meta-knowledge transfer for inductive knowledge graph embedding. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 927–937.

[8] Ali Farhadi, Ian Endres, Derek Hoiem, and David Forsyth. 2009. Describing objects by their attributes. In *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 1778–1785.

[9] Andrea Frome, Greg S Corrado, Jon Shlens, Samy Bengio, Jeff Dean, Marc'Aurelio Ranzato, and Tomas Mikolov. 2013. Devise: A deep visual-semantic embedding model. *Advances in neural information processing systems* 26 (2013).

[10] Zhenyong Fu, Tao Xiang, Elyor Kodirov, and Shaogang Gong. 2015. Zero-shot object recognition by semantic manifold distance. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2635–2644.

[11] Junyu Gao, Tianzhu Zhang, and Changsheng Xu. 2019. I know the relationships: Zero-shot action recognition via two-stream graph convolutional networks and knowledge graphs. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 8303–8311.

[12] Yuxia Geng, Jiaoyan Chen, Zhuo Chen, Jeff Z Pan, Zhiquan Ye, Zonggang Yuan, Yantao Jia, and Huajun Chen. 2021. Ontozsl: Ontology-enhanced zero-shot learning. In *Proceedings of the Web Conference 2021*. 3325–3336.

[13] Yuxia Geng, Jiaoyan Chen, Zhiquan Ye, Zonggang Yuan, Wei Zhang, and Huajun Chen. 2021. Explainable zero-shot learning via attentive graph convolutional network and knowledge graphs. *Semantic Web* 12, 5 (2021), 741–765.

[14] Yuxia Geng, Jiaoyan Chen, Wen Zhang, Yajing Xu, Zhuo Chen, Jeff Z. Pan, Yufeng Huang, Feiyu Xiong, and Huajun Chen. 2022. Disentangled ontology embedding for zero-shot learning. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*. 443–453.

[15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).

[16] Takuo Hamaguchi, Hidekazu Oiwa, Masashi Shimbo, and Yuji Matsumoto. 2017. Knowledge transfer for out-of-knowledge-base entities: A graph neural network approach. *arXiv preprint arXiv:1706.05674* (2017).

[17] Junji Jiang, Hongke Zhao, Ming He, Likang Wu, Kai Zhang, and Jianping Fan. 2023. Knowledge-Aware Cross-Semantic Alignment for Domain-Level Zero-Shot Recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 965–975.

[18] Michael Kampffmeyer, Yinbo Chen, Xiaodan Liang, Hao Wang, Yujia Zhang, and Eric P Xing. 2019. Rethinking knowledge graph propagation for zero-shot learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 11487–11496.

[19] Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[20] Elyor Kodirov, Tao Xiang, and Shaogang Gong. 2017. Semantic autoencoder for zero-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3174–3183.

[21] Xiaoyu Kou, Yankai Lin, Shaobo Liu, Peng Li, Jie Zhou, and Yan Zhang. 2020. Disentangle-based Continual Graph Representation Learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2961–2972.

[22] Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. 2013. Attribute-based classification for zero-shot visual object categorization. *IEEE transactions on pattern analysis and machine intelligence* 36, 3 (2013), 453–465.

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

[24] Jingjing Li, Mengmeng Jing, Ke Lu, Zhengming Ding, Lei Zhu, and Zi Huang. 2019. Leveraging the invariant side of generative zero-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7402–7411.

[25] Lu Liu, Tianyi Zhou, Guodong Long, Jing Jiang, and Chengqi Zhang. 2020. Attribute propagation network for graph zero-shot learning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 4868–4875.

[26] James MacQueen et al. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1. Oakland, CA, USA, 281–297.

[27] 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* (2013).

[28] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

[29] Ruizhi Qiao, Lingqiao Liu, Chunhua Shen, and Anton Van Den Hengel. 2016. Less is more: zero-shot learning from online textual documents with noise suppression. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2249–2257.

[30] 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 *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 8673–8680.

[31] Abhinaba Roy, Deepanway Ghosal, Erik Cambria, Navonil Majumder, Rada Mihalcea, and Soujanya Poria. 2022. Improving zero-shot learning baselines with commonsense knowledge. *Cognitive Computation* 14, 6 (2022), 2212–2222.

[32] Haseeb Shah, Johannes Villmow, Adrian Ulges, Ulrich Schwanecke, and Faisal Shafait. 2019. An open-world extension to knowledge graph completion models. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 3044–3051.

[33] Baoxu Shi and Tim Weninger. 2018. Open-world knowledge graph completion. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 32.

[34] Robert L Solso, M Kimberly MacLin, and Otto H MacLin. 2005. *Cognitive psychology*. Pearson Education New Zealand.

[35] Komal Teru, Etienne Denis, and Will Hamilton. 2020. Inductive relation prediction by subgraph reasoning. In *International Conference on Machine Learning*. PMLR, 9448–9457.

[36] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. 2016. Complex embeddings for simple link prediction. In *International conference on machine learning*. PMLR, 2071–2080.

[37] Jiahui Wang, Likang Wu, Hongke Zhao, and Ning Jia. 2023. Multi-view enhanced zero-shot node classification. *Information Processing & Management* 60, 6 (2023), 103479.

[38] Peifeng Wang, Jialong Han, Chenliang Li, and Rong Pan. 2019. Logic attention based neighborhood aggregation for inductive knowledge graph embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 7152–7159.

[39] Wei Wang, Vincent W Zheng, Han Yu, and Chunyan Miao. 2019. A survey of zero-shot learning: Settings, methods, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)* 10, 2 (2019), 1–37.

[40] Xuesong Wang, Chen Chen, Yuhu Cheng, and Z Jane Wang. 2016. Zero-shot image classification based on deep feature extraction. *IEEE Transactions on Cognitive and Developmental Systems* 10, 2 (2016), 432–444.

[41] Xin Wang, Jiawei Wu, Da Zhang, Yu Su, and William Yang Wang. 2019. Learning to compose topic-aware mixture of experts for zero-shot video captioning. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 33. 8965–8972.

[42] Xiaolong Wang, Yufei Ye, and Abhinav Gupta. 2018. Zero-shot recognition via semantic embeddings and knowledge graphs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 6857–6866.

[43] Likang Wu, Junji Jiang, Hongke Zhao, Hao Wang, Defu Lian, Mengdi Zhang, and Enhong Chen. 2023. KMF: knowledge-aware multi-faceted representation learning for zero-shot node classification. *arXiv preprint arXiv:2308.08563* (2023).

[44] Yongqin Xian, Tobias Lorenz, Bernt Schiele, and Zeynep Akata. 2018. Feature generating networks for zero-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5542–5551.

[45] Wenjia Xu, Yongqin Xian, Jiuniu Wang, Bernt Schiele, and Zeynep Akata. 2020. Attribute prototype network for zero-shot learning. *Advances in Neural Information Processing Systems* 33 (2020), 21969–21980.

[46] 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* (2014).

[47] Kai Zhang, Qi Liu, Zhenya Huang, Mingyue Cheng, Kun Zhang, Mengdi Zhang, Wei Wu, and Enhong Chen. 2022. Graph adaptive semantic transfer for cross-domain sentiment classification. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1566–1576.

[48] Kai Zhang, Qi Liu, Hao Qian, Biao Xiang, Qing Cui, Jun Zhou, and Enhong Chen. 2021. EATN: An efficient adaptive transfer network for aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering* 35, 1 (2021), 377–389.

[49] Kai Zhang, Hefu Zhang, Qi Liu, Hongke Zhao, Hengshu Zhu, and Enhong Chen. 2019. Interactive attention transfer network for cross-domain sentiment classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 5773–5780.

[50] Li Zhang, Tao Xiang, and Shaogang Gong. 2017. Learning a deep embedding model for zero-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2021–2030.

[51] Hongke Zhao, Chuang Zhao, Xi Zhang, Nanlin Liu, Hengshu Zhu, Qi Liu, and Hui Xiong. 2023. An ensemble learning approach with gradient resampling for class-imbalance problems. *INFORMS Journal on Computing* 35, 4 (2023), 747–763.

[52] Ming Zhao, Weijia Jia, and Yusheng Huang. 2020. Attention-based aggregation graph networks for knowledge graph information transfer. In *Advances in Knowledge Discovery and Data Mining: 24th Pacific-Asia Conference, PAKDD 2020, Singapore, May 11–14, 2020, Proceedings, Part II 24*. Springer, 542–554.

[53] Yizhe Zhu, Mohamed Elhoseiny, Bingchen Liu, Xi Peng, and Ahmed Elgammal. 2018. A generative adversarial approach for zero-shot learning from noisy texts. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1004–1013.

[54] Yizhe Zhu, Jianwen Xie, Bingchen Liu, and Ahmed Elgammal. 2019. Learning feature-to-feature translator by alternating back-propagation for generative zero-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 9844–9854.