Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 6300–6309
Marseille, 20-25 June 2022

Challenging the Assumption of Structure-based embeddings in Few- and Zero-shot Knowledge Graph Completion

Filip Cornell[1,3], **Chenda Zhang**[2], **Šarūnas Girdzijauskas**[1], **Jussi Karlgren**[3]
[1]KTH Royal Institute of Technology, [2]Carnegie Mellon University, [3]Gavagai
fcornell@kth.se, chendaz@andrew.cmu.edu, sarunasg@kth.se, jussi@lingvi.st

Abstract

In this paper, we report experiments on Few- and Zero-shot Knowledge Graph completion, where the objective is to add missing relational links between entities into an existing Knowledge Graph with few or no previous examples of the relation in question. While previous work has used pre-trained embeddings based on the structure of the graph as input for a neural network, nobody has, to the best of our knowledge, addressed the task by only using textual descriptive data associated with the entities and relations, much since current standard benchmark data sets lack such information. We therefore enrich the benchmark data sets for these tasks by collecting textual description data to provide a new resource for future research to bridge the gap between structural and textual Knowledge Graph completion. Our results show that we can improve the results for Knowledge Graph completion for both Few- and Zero-shot scenarios with up to a two-fold increase of all metrics in the Zero-shot setting. From a more general perspective, our experiments demonstrate the value of using textual resources to enrich more formal representations of human knowledge and in the utility of transfer learning from textual data and text collections to enrich and maintain knowledge resources.

Keywords: Knowledge Graph completion, Meta-learning, Zero-shot learning, textual enrichment

1. Introduction

Knowledge graphs are formalized representations of world knowledge, useful in many tasks such as Question Answering, Language Representation Learning and Recommender Systems (Ji et al., 2021). They consist of nodes representing *entities* of interest and directed edges representing different *relations* between the entities. A knowledge graph is frequently incomplete: not every relation has been observed at the time the its construction (Xiong et al., 2018; Min et al., 2013). This is the basis for the *Knowledge Graph completion* task: to infer missing links, given a graph. More formally, Knowledge Graphs can be described as a set of *triplets* $(h, r, t) \in \mathcal{T}$ with a *head* entity h, a *tail* entity t, and the relationship r defining their connection. A Knowledge Graph can therefore be described as $\mathcal{KG} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$, where $\mathcal{E}$ denotes the set of entities and $\mathcal{R}$ is the set of relationship types.

Knowledge Graph completion is mostly addressed by scoring candidate triplets (h, r, t) for relation r by some assigned validity score. True or valid triplet candidates rank higher by the validity score and false ones lower. Previous work (Bordes et al., 2013; Yang et al., 2015; Dettmers et al., 2018; Sun et al., 2019) have moved a Knowledge Graph to a Euclidean vector space, assigning each entity and relation a d-dimensional vector. These vectors are then trained, given a scoring function $f(\cdot)$, to either maximize the score $f(h, r, t)$ for true triplets and minimize the score $f(h, r, t)$ for false ones. Examples of these include **TransE** (Bordes et al., 2013) and **DistMult** (Yang et al., 2015) and we refer to this type of Knowledge Graph embeddings as *structure-based*, as they fully and solely rely on the structure of the Knowledge Graph.

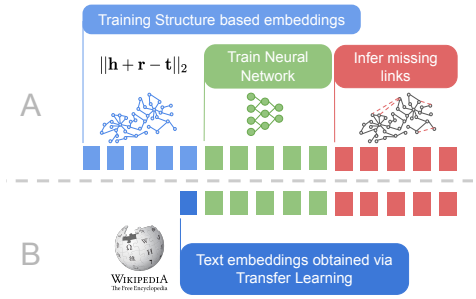

Figure 1: **A:** Procedure of previous approaches in Few- and Zero-shot Knowledge Graph completion. **B:** Instead of training structural embeddings, we obtain representations through textual descriptors, reducing the overall effort.

However, as pointed out by Xiong et al. (Xiong et al., 2018), structural Knowledge Graph embeddings often fall short in predicting relations seldom observed in the graph. To alleviate this issue, previous works have framed the problem in a *Few- or Zero-shot* learning scenario, gaining considerable improvements in completing links for rare relations. To the best of our knowledge, no one has yet successfully, in a few- or zero-shot setting, addressed this task by beginning only from the textual descriptions available in the actual Knowledge Graph: instead, previous efforts have focussed on mainly leveraging the structure of the graph. Our work enriches two Knowledge Graph benchmark data sets with texts associated with them and uses those texts for inferring missing relations with greatly improved results compared to a structural representation.

Few-shot and Zero-shot learning are learning frame-

works where an existing pre-trained representation is quickly adapted to new tasks with a very small amount of previously seen examples, or none at all. They both rely on the existing representation having been previously trained on suitably large and general data, and then leverage this external knowledge to predict for the unseen or barely seen cases (Chen et al., 2021a).

Previous approaches for Few- and Zero-shot Knowledge Graph completion have used a *transfer learning* approach by first leveraging *structural* pre-trained Knowledge Graph embeddings to embed incoming triplets to the network trained on an existing Knowledge Graph (see Figure 1A). While this has proven successful, it is limited to the information derived from the structure of the graph and requires explicit and costly training on the Knowledge Graph at hand. This pre-training process requires careful consideration during training (Ruffinelli et al., 2019). Word embedding representations such as Word2Vec or GloVe, and recent transformer-based models such as BERT (Mikolov et al., 2013; Pennington et al., 2014; Devlin et al., 2019), on the other hand, are usually trained on larger corpora of knowledge, and promise to be useful if combined with an appropriate meta-learning architecture. Additionally, zero- and few-shot knowledge graph completion methods today inherently capture the structure of the graph by design, which allows us to simultaneously leverage structural and textual information by starting with textual features.

In this paper we represent the entities h, t and relations r of a Knowledge Graph by using word embeddings trained on general textual resources to allow us to infer the validity of suggested (h, r, t) triplets through their textual descriptions. This allows us to capture relationships that have little support in structural examples used by previous approaches. Using previously published benchmarks, enriched by us with textual labels, we experimentally demonstrate an up to a two-fold increase in performance.

2. Data sets

For these experiments, we use benchmarks based on two well-established knowledge graphs: NELL (Mitchell et al., 2018) and Wikidata (Vrandečić and Krötzsch, 2014) and the Wikidata benchmarks Wiki-ZS (Qin et al., 2020) and Wiki-One (Xiong et al., 2018).

The benchmark data sets Wiki-One and Wiki-ZS do not include textual descriptions of entities and Wiki-One does not have textual descriptions for relations. As part of our experimentation, we therefore enrich these benchmark sets. We collect textual descriptions[1] for every entity in the Wiki-One and Wiki-ZS data sets and for the relations in the Wiki-One data set and add more relational descriptions for the NELL-One and Wiki-One data sets.

Wiki-One and Wiki-ZS Wiki-One and Wiki-ZS are both subsets of the Wikidata knowledge base (Vrandečić and Krötzsch, 2014). This data set is significantly larger, as they both are subsets from a Wikidata dump from 2018. Both relations and entities are defined by their ID, a string starting with Q for entities and P for relations, followed by a number.

For NELL-ZS and Wiki-ZS, the test, validation, and training sets for Knowledge Graph completion benchmarks have thorough textual descriptions collected by Qin et al (Qin et al., 2020), but they lack textual features for all entities. However, the relations included in Wiki-One but excluded from Wiki-ZS have no textual descriptions. Our work completes these data sets for the entities and relations without text, and by doing so, we broaden the type of methods applicable to these data sets.

Expanding Wiki-One and Wiki-ZS In the Wiki data sets, the entities in the data set are only represented by their IDs. To enrich the data set, we collect three types of textual information for the entities and relations from the Wikidata knowledge base on which the data sets are based.

- `label`: The label is usually the name of the entity, such as a person's or a location's name.
- `description`: A brief textual description of an entity which can be in any of several languages.
- For entities, we also collect the attribute `instance of`, described as *"that class of which this subject is a particular example and member"*[2].

Some examples of textually enriched entities and relations are shown in Table 2. The descriptors were downloaded for all existing entities and relations contained in the Wiki-One and Wiki-ZS data sets. The knowledge base on which these data sets are based evolves continuously, illustrating the necessity for updating representations: out of the $4\,838\,244$ entities in the data set snapshot, $16\,779$ ($\approx 0.35\,\%$) had been deleted from the Wikidata Knowledge Base and thus lack descriptive labels and texts. These entities were labeled with "*Unknown*". Besides the deleted entities, $404\,298$ entities ($\approx 8\,\%$) lacked labels in any language, $307\,649$ lacked description texts, and 2697 had no superclass *instance* attribute as shown in Table 3.

These collected data were first split into two sections: one with entities and relations where both English labels and description texts were present and another section where entities and relations lack either a label or a description text in English.

For entities in the latter section, we included every available description and label from thirteen languages: English, French, Spanish, Italian, Portuguese, Chinese,

[1] Available at: `https://bit.ly/3MHYCxV`

[2] Described at `https://www.wikidata.org/wiki/Property:P31`

Setting	Data set	# Ent.	# Rel.	# Triplets	Splits	Word count
Few-shot	NELL-One	68,545	358	181,209	51/5/11	3.3 ± 1.2
	Wiki-One	4,838,244	822	5,829,240	133/16/34	8.6 ± 5.1
Zero-shot	NELL-ZS	65,567	181	188,392	139/10/32	3.3 ± 1.1
	Wiki-ZS	605,812	537	724,967	469/20/48	10.3 ± 5.5

Table 1: Data sets used in the experiments. Word count denotes the number of words per entity and relation on average, with its standard deviation. Split denotes the number of relations in each pre-defined split (train, validation and test) of the data sets.

data set	Data type	ID	Label	Description	Instance
Wiki	Entities	Q8180	Feyzin disaster	fire in a refinery near the town of Feyzin, France	desastre
		Q83115	Fantasy World Dizzy	1989 video game	video game
		Q4776245	Antonia Juhasz	American journalist	human
		Q17528437	Church of St Mary	historic church in Ellingham, Hampshire, England, United Kingdom	church
	Relations	P37	official language	language designated as official by this item	N/A
		P129	physically interacts with	physical entity that the subject interacts with	N/A
		P277	programming language	the programming language(s) in which the software is developed	N/A
		P2632	place of detention	place where this person is or was detained	N/A
NELL	Entities	concept:company:kmau	kmau	N/A	company
		concept:male:ability	ability	N/A	male
	Relations	concept:colorofobject	color of object	Object has color (NEIL)	N/A
		concept:lakeinstate	lake in state	The state/province a river is in.	N/A

Table 2: Examples of the entities and relations for both the Wiki and NELL benchmarks, with their textual enrichment data.

Japanese, Swedish, Arabic, Korean, Dutch, Russian, and Indonesian. We then translated the missing texts into English from the thirteen different languages using the MarianMT models (Junczys-Dowmunt et al., 2018) seen in Table 4. In addition, out of 29 184 unique "instance" attributes, 189 did not have English names. For these, we used the Google translate API to translate them into English.

The relations were far fewer, and only five of them lacked English language descriptions. For P2329, P3592 and P3494 we translated the description text to English manually. Relations P2157 and P2439 had been removed from Wikidata, but their information was recovered using the Wikidata log[3].

	Total	Deleted	No label	No description	No instance
Entities	4,838,244	16,779	404,298	307,649	2,697
Relations	822	2	0	0	N/A

Table 3: Deleted and missing entries in the data sets collected from the Wikidata knowledge base.

NELL-One and NELL-ZS NELL (Mitchell et al., 2018) is a Knowledge Graph initially constructed as an attempt to structure the knowledge of the internet. Both data sets consist of about 65 000 entities and a few hundred relations, with each entity and relation defined by an ID which constitutes a descriptive phrase which defines the concept it refers to. These ID:s have been segmented into words by us, provided in the enriched data set resource. While we retrieved longer tex-

[3] https://www.wikidata.org/wiki/Special:Log

Language	MarianMT model	# Labels	# Descriptions
Arabic	opus-mt-ar-en	22,082	93,148
Chinese	opus-mt-zh-en	150,654	1,591
Dutch	opus-mt-nl-en	73,938	151,658
French	opus-mt-roa-en	75,632	65,338
German	opus-mt-de-en	59,371	158,969
Indonesian	opus-mt-id-en	1	113
Italian	opus-mt-roa-en	12,856	9,870
Japanese	opus-mt-ja-en	13,971	7,791
Korean	opus-mt-ko-en	314	63
Russian	opus-mt-ru-en	141,800	26,520
Portuguese	opus-mt-roa-en	7,284	1,409
Spanish	opus-mt-roa-en	23,783	13,188
Swedish	opus-mt-sv-en	76,099	30,172
Total		657,785	559,830

Table 4: The different languages from which the descriptive texts were translated, and the number of items translated to English from each language.

tual descriptions for the relations through their definitions from the NELL knowledge base through its metadata and therefore extended the NELL data sets for our tasks (Mitchell et al., 2018), we were not able to obtain longer textual descriptions for the entities. This gives us less informative textual features for this data set, but provides us with a comparison between scenarios where textual descriptions are sparse and where they are verbose, as in the Wiki-One and Wiki-ZS cases. For examples of entities from the NELL graph, we refer to Table 2. For entities, we use the instance relation and the definition phrase as the label; for relations only the definition phrase is used, again as the label.

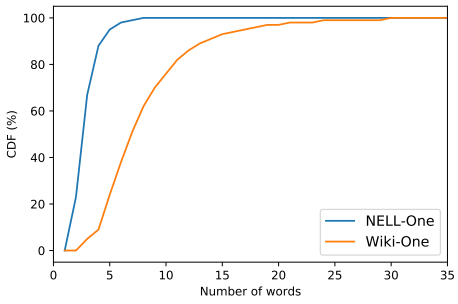

Figure 2: Cumulative distributions of the number of words in the Wiki-One and NELL-One texts.

3. Related work

Few-shot Learning Few-shot learning, also known as meta learning (Huisman et al., 2021), has the objective to quickly adapt a representation to new tasks barely seen in the data before. This is done by using a support set $S_{\mathcal{T}}$ and a query set $\mathcal{Q}_{\mathcal{T}}$ for every task, and the support set size is usually very small, usually of a size 1 to 5. In a Knowledge Graph completion scenario, a task $\mathcal{T}$ involves predicting the correctness of triples for some relationship t. Using a support set of true examples $\mathcal{S}_r = \{(h, r, t) \mid h, t \in \mathcal{E}, r \in \mathcal{R}_{\text{few}}\}$, and a set of queries $\mathcal{Q}_r = \{(h, r, t) \mid h, t \in \mathcal{E}, r \in \mathcal{R}_{\text{few}}\}$, the objective is to predict whether the query triples are true or not. By moving to a K-shot scenario using K examples in every iteration, one optimally leverages the examples obtained previously for a specific task by using a neural meta-learning approach. These are either similarity-based, such as Matching or Prototypical Networks (Vinyals et al., 2016; Snell et al., 2017) or optimization-based, like Model-Agnostic Meta Learning (MAML) (Finn et al., 2017).

Zero-shot Learning Zero-shot learning refers to the task of predicting classes and tasks previously not seen at all (Geng et al., 2021; Chen et al., 2021a). When the Zero-shot learning is Knowledge-based (as is in our case) one uses some auxiliary information to align the task at hand and enable a neural model to predict the unseen classes.

Knowledge Graph Completion Most Knowledge Graph completion methods are structural, i.e., completely relying on the structure of the Knowledge Graph, and move the entities and relations into either a joint or two separate Euclidean spaces $\mathbb{R}^D$ by using the information given by the graph's structure. For many Knowledge Graph embedding methods, a set of true (h, r, t) and false triplets (h, r, t'), are sampled in each iteration. A contrastive loss is then used as in (1) to maximise the score for true triplets and minimize it for false ones. In both Zero- and Few-shot Knowledge Graph Completion, this has been used as the main objective function.

$$\mathcal{L} = \sum_{(h,r,t)} \sum_{(h,r,t')} \gamma + \text{score}_{(h,r,t')} - \text{score}_{(h,r,t)} \quad (1)$$

Two examples of structural embeddings are **TransE** (Bordes et al., 2013) and **DistMult** (Yang et al., 2015). While TransE is optimized to be additive such that the embedding of the head and the relation added approximates the tail through addition, such that $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$, DistMult is optimized to translate in a multiplicative such that $\mathbf{h} \cdot \mathbf{r} \approx \mathbf{t}$.
Several previous efforts in Knowledge Graph completion have demonstrated the effectiveness of exploiting textual information (Hu et al., 2021; Xiao et al., 2017; Xie et al., 2016; Socher et al., 2013). Xiao et al. (Xiao et al., 2017) projected the Knowledge Graph embedding space onto the word embedding space by using textual features, and Xie et al. (Xie et al., 2016) used both the structure and the descriptions of the entities to create a model capable of Zero-shot learning. The model for Knowledge Graph embeddings by Xie et al was built jointly from both structural and textual descriptions and proved useful in a Zero shot context as well. This shows how a richer representation of Knowledge Graph content is useful in a cold-start or a Zero-shot situation. Our experiments are designed to show how this holds for relation triple inference as well, specifically for unseen relations.

Setting In both Few- and Zero-shot Knowledge Graph Completion, we consider the setting of having a background knowledge graph $\mathcal{BG} : \{(h, r, t) \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}\}$ where $\mathcal{E}$ is the set of all entities and $\mathcal{R}$ is the set of common relations. The task is to predict likely triplets for a set of rare or unseen relations $\mathcal{R}_{\text{few}} = \{r_1, .., r_{|\mathcal{R}_{\text{few}}|} | \mathcal{R}_{\text{few}} \cap \mathcal{R} = \emptyset\}$. In each training iteration, a correct triplet $(h, r, t), r \in \mathcal{R}_{\text{few}}$ and a query triplet (h', r, t') are sampled, where R_r is a set of true triplets for a task relation is used as a support example. In few-shot learning, a support set $\mathcal{S}_r$ is also sampled.

Few-shot Knowledge Graph completion Few-shot Knowledge Graph completion has been done both through similarity-based (Xiong et al., 2018; Zhang et al., 2020; Sheng et al., 2020) and MAML-based approaches (Chen et al., 2019; Lv et al., 2019). One of the first published studies is GMatching (Xiong et

al., 2018) in a one-shot relational learning setting in which the one-hop neighborhood graph around each head and tail entity was aggregated by summing up embeddings of every entity and relation within one hop from the entity. The query and support triplet were then inserted into an LSTM-based Matching Network (Vinyals et al., 2016), producing a score reflecting the likelihood of the query. A similar work, improving and extending upon GMatching was FSRL (Zhang et al., 2020), by using attention mechanisms to weight neighbors differently. However, these attention vectors were static for an entity and did not change over relations. As in the example originally given by (Sheng et al., 2020), the entity `Microsoft` might be more important than `MelindaGates` when predicting for task relation `CeoOf`, and the opposite might be true in the case of `FatherOf`. (Sheng et al., 2020) therefore introduced the Adaptive Attentional Network (FAAN) which we explain in more detail in Section 4.2.

Zero-shot Knowledge Graph completion One of the first published studies moving Knowledge Graph completion to a zero-shot setting was DKRL (Xie et al., 2016), a deep convolutional neural network. That work shows that using textual descriptions combined with a structural embedding approach a deep Convolutional Neural Network (CNN) is able to learn to complete triplets for unseen entities.

Our experiments focus on Zero-shot learning where the *relations* to be inferred have not previously appeared in the Knowledge Graph under consideration. Some previous efforts in this direction include the Zero-shot Generative Adversarial Network (ZS-GAN) (Qin et al., 2020) and the Ontological Zero-shot Learning (OntoZSL) model (Geng et al., 2021). Both of these rely on generating plausible embeddings for previously unobserved relations using a Generative Adversarial Network (GAN) (Goodfellow et al., 2014), given a set of descriptions of those unseen relations. These approaches both make use of auxiliary information about these relations to be able to infer triplets with them. We use both ZS-GAN and OntoZSL in our experiments below. Another work related to ours is the work by Li et al (Li et al., 2020) which uses logic-guided learning for relation classification, i.e., identifying the correct relation for a head, relation and tail. Li et al. compare word embeddings and structural embeddings and conclude that structural embeddings give better results than word embeddings. This contradicts our findings, but there are a few distinctive differences. Firstly, they train their word embeddings only on the data set they construct, and therefore limit the amount of knowledge encoded. Secondly, they do not make use of contextualised transformer-based embeddings, and limit themselves to using DeVISE (Frome et al., 2013) and ConSE (Norouzi et al., 2013), two models originally created for image-related zero-shot classification rather than Knowledge Graph relation classification.

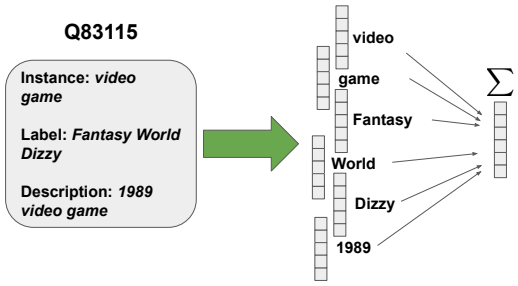

Figure 3: Example of an entity from the Wikidata knowledge base. With SIF, we embed the entities and relations using a weighting of the words it has.

Integrating textual features Numerous studies have used textual information to infer missing links in Knowledge Graphs such as mentioned DKRL and OWL2Vec (Chen et al., 2021b) which embeds the nodes in a Knowledge Graph using a random walk-based approach, while simultaneously incorporating both logical, structural, and textual information. Other studies (Wang et al., 2021) have developed large-scale language models such as KEPLER where the model is completely reliant on text, requiring a larger corpora of knowledge for every entity to be embedded and proving powerful inductive capabilities. Our focus is to add a data set with a smaller amount of text, and show that this can still be highly useful.

4. Method

4.1. Embedding entities and relations through text

The first step in our approach is to convert the textual descriptions we have collected into vector-based representations. In our experiments, we use two methods of embedding the entities and relations: Smooth Inverse Frequency embeddings (SIF) and transformer-based sentence embeddings (BERT).

Transformer-based sentence embeddings In our Zero-shot learning experiments, we embed the text collected using a pre-trained sentence embedder `paraphrase-MiniLM-L12-v2`, producing semantic sentence embeddings of 384-dimensions (Reimers and Gurevych, 2019). Initially, these are trained on sentences. A high cosine similarity between two sentences indicate that they have the same semantic meaning, making it suitable to represent a collection of information. However, we found in our experiments that we were not able to utilize the sentence embedding together with **FAAN** due to scalability constraints in terms of memory resource limits. We therefore employ another method as well, known as Smooth Inverse Frequency (Arora et al., 2019) which leverages word embeddings and are less computationally costly, much due to their lower dimensionality.

Smooth Inverse Frequency embeddings SIF is a method that produces compact representations for sets

of words using precomputed word vectors such as Word2Vec (Mikolov et al., 2013) or GloVe (Pennington et al., 2014) weighted by the observed frequencies of word occurrence in a language. The representation uses a Singular Value Decomposition computation step to modulate the effect of frequent phrases and words by removing the first N principal components in a matrix of generated vectors.

SIF embeddings are appropriate to test for our experiments for a few reasons. Firstly, SIF is fast and relatively cheap compared to other approaches, assuming access to pre-trained word embeddings. SIF vectors are low-dimensional in comparison to transformer-based models such as BERT (Devlin et al., 2019), and of similar dimensionality as other Knowledge Graph embedding techniques. Secondly, previous work (Allen et al., 2021) has also shown a close relationship between Knowledge Graph embeddings and context-free word embeddings, as they and TransE share their translational characteristics (Mikolov et al., 2013; Bordes et al., 2013). Therefore, SIF is a well-suited replacement for TransE since we remain in the word embedding space, maintaining these properties. Thirdly, SIF does not rely on word order or local frequency of occurrence, which makes it suitable for analysis of brief textual material such as the ones at hand, as the vast majority of descriptions contain less than 15 words (see 2). This allows us to concatenate the labels, descriptions, and superclass instance relations into simple sets of words (see Figure 3). In the singular value decomposition of SIF, we remove the first five principal components and use pre-trained GloVe vectors[4] `glove-wiki-gigaword-100` in 100 dimensions.

4.2. Few-shot setting

In the Few-shot setting, we investigate whether the Adaptive Attentional Network (FAAN) by Sheng et al. (Sheng et al., 2020) can benefit from the textual embeddings derived from the data collected.

FAAN FAAN is, as mentioned previously, an attention-based neural network which maximizes the scores for true triplets and minimizes the score for false ones. In each iteration, a set of K support examples for a rare relation $r \in R$, $\mathcal{S}_r = \{(h_i, r, t_i) \mid i \in 1..K, r \in R\}$ are sampled along with a true query $Q = (h_Q, r, t_Q)$ and a false example $Q_f = (h_{Q_f}, r, t_{Q_f})$. For each entity in all triplets involved, the pre-trained vector representations of the neighborhood on an entity $\mathcal{N}_e = \{(r_{nbr}, e_{nbr} | (h, r_{nbr}, e_{nbr})\}$ are aggregated into a representation of each entity $\mathbf{e} \in \mathcal{S}_r \cup Q \cup Q_f$ through attention-based mechanisms. For each triplet, the relation r is given an estimate $\mathbf{e}_r = \mathbf{e}_t - \mathbf{e}_h$, based on the translational property of TransE. Each triplet's set of head, relation and tail vectors $(\mathbf{e}_h, \mathbf{e}_r, \mathbf{e}_t)$ are then concatenated and passed onto an transformer encoder, where each triplet is transformed into a single representation $\mathbf{z}_{(h,r,t)}$. The vectors $\mathbf{z}_{(h,r,t)}$ of the support set triplets are then aggregated into a representation $\mathbf{S}_r$, adapted to the query $\mathbf{q}_r$, again through attention. Finally, a similarity score $\phi(\mathbf{Q}_r, \mathbf{S}_r)$ is given by their dot product, trained to be maximized if $\mathbf{Q}_r$ is a true triplet and minimized otherwise. For a more in-depth explanation, we refer to Sheng et al. (Sheng et al., 2020).

[4]Available at https://nlp.stanford.edu/projects/glove/.

4.3. Zero-shot setting

In a Zero-shot setting, we investigate the benefit of using textual embeddings using two relevant previous works: the **ZS-GAN** (Qin et al., 2020) and **OntoZSL** (Geng et al., 2021).

ZS-GAN The ZS-GAN, introduced by Qin et al. (Qin et al., 2020) is a zero-shot architecture consisting of three components; the *feature encoder*, the *generator*, and the *discriminator*. As a first step, a feature encoder is trained to learn a better distribution of the data from the pre-trained embeddings for entity pairs. It consists of two encoders; one considering the head and the tail jointly ($f_{\text{pair}}(v_h, v_r)$ as in Equation (2)), and a neighbor encoder $f_{\text{N}}(\mathcal{N}_e)$ (see Equation (3)) taking the neighbors of an entity into account. Here, σ denotes the tanh-function, $W_{\text{pair}} \in \mathbb{R}^{d \times 2d}$ is a trainable weight matrix and $b \in \mathbb{R}^d$ its bias. A representation of a triplet $v_{(h,t)}$ is then concatenated as in (4), producing the final representation from the feature encoder. The feature encoder is pre-trained through a contrastive loss where true and false pairs are sampled, and the cosine similarity between the outputted representation $v_{(h,t)}$ and the original triplet (h, r, t) is maximized for the true examples and minimized for the false.

$$f_{\text{pair}}(v_h, v_r) = \sigma(W_{\text{pair}}(v_h \oplus v_t)) + b_{\mathcal{N}} \quad (2)$$

$$f_{\text{N}}(\mathcal{N}_e) = \sigma(\frac{1}{|\mathcal{N}_e|} \sum_{(r,e) \in \mathcal{N}_e} W(v_r \oplus v_e) + b) \quad (3)$$

$$v_{(h,t)} = f_{\text{N}}(\mathcal{N}_h) \oplus f_{\text{pair}}(v_h, v_t) \oplus f_{\text{N}}(\mathcal{N}_t) \quad (4)$$

The generator and the discriminator are trained with a pre-trained feature encoder. A textual representation of an unseen relation $v_{r_{\text{text}}}$, produced by word embeddings weighted by their *tf.idf*-score (Jones, 1972) is input into the generator $f_{Gen}(v_{r_{\text{text}}} \oplus z)$, concatenated with normally distributed noise $z \sim \mathcal{N}(0, 1)$. The generator $f_{Gen}(\cdot)$, which is a two-layer feed-forward neural network, then outputs a representation which the discriminator $f_{Disc}(\cdot)$ learns to classify as true or false. The discriminator is fed both outputs from the feature encoder and false examples generated by the generator and learns to produce plausible embeddings for the relations. These can then be used to complete the missing links for new, unseen relation types even if the relation has not been seen previously. For further reading and implementation details, we refer to (Qin et al., 2020).

OntoZSL OntoZSL (Geng et al., 2021) improves upon ZS-GAN by incorporating ontological information. It bears much resemblance to the ZS-GAN in terms of its architecture, but with one additional module: the **Ontology Encoder**. The Ontology Encoder takes an ontological schema of a Knowledge Graph, encodes it as an embedding, and inputs it into the generator to generate a plausible structure- and text aware embedding which contains richer information than a vector based only on a textual description of the relation. Apart from incorporating ontological information, the neural architecture of OntoZSL to synthesize examples of unseen relation types is similar to that of ZS-GAN. Entities are passed along with neighbors to a *Feature Encoder*, which is then passed to a *Discriminator* as a true example. For further details, we refer to Geng et al. (Geng et al., 2021).

5. Experiments

In our experiments, which we make publicly available (Cornell, 2022), our goal is to show that our textual descriptions can easily replace the structure-based embeddings in the three existing neural architectures introduced above and show the benefit of using the data we collected for enriching the data sets for this task.

5.1. Few-shot experiments

We run experiments on the NELL-One and Wiki-One data sets and perform five-shot learning using the **FAAN** architecture. We use the same hyperparameters on both data sets as the original authors (Sheng et al., 2020); for more details, we refer to our experiments (Cornell, 2022). We report four metrics: the hit ratios Hits@10, Hits@5, and Hits@1, giving the ratio of correct entities landing in the top X results, and the Mean Reciprocal Rank (MRR), giving the average of the inverse rank as in (5). We set X to be 1, 5, and 10.

$$\text{MRR} = \frac{1}{N}\sum_{i=1}^{n}\frac{1}{\text{rank}_i} \tag{5}$$

For both NELL-One and Wiki-One, we use 100-dimensional SIF embeddings and run experiments with TransE, DistMult, and SIF with the same settings. We also conduct ablation studies and remove each text attribute one at a time to see which contribute most when using SIF with the 100-dimensional GloVe vectors. This helps us understand further what type of information in the word embeddings that are important. As mentioned in Section 4.1, we are due to memory constraints limited to SIF, and cannot perform experiments using the transformer-based embeddings.

5.2. Zero-shot experiments

In the Zero-shot experimental setting, we re-use the **ZS-GAN** and **OntoZSL** architectures. As in the Few-shot experiments, we use **TransE** and **DistMult** Knowledge Graph embeddings, and also replace these with SIF and BERT-based embeddings. We again use the same hyperparameters used in the original experiments (Cornell, 2022) and compare under similar settings. We re-use parts of the code uploaded by the authors of ZS-GAN (Qin et al., 2020) and OntoZSL (Geng et al., 2021), respectively.

6. Results

Few-shot results Table 6 displays our results. For NELL-One, the improvement from using SIF is not as great as for Wiki-One. We hypothesize that the discrepancy is due to the sparser text descriptions in NELL, in particular for the entities.

The ablation study in Table 7 shows that the labels carry most of the information in both data sets, which indicates that information about entity names is captured through word vectors trained on large text corpora. Using `instance` seems to make little difference on Wiki-One. We believe this is because the descriptive text in `instance` often is terse, generic, and unspecific: there are only 29 184 unique `instance` descriptions in the Wiki-One data set and e.g. approximately 33 % of all entities have the label `human`.

Zero-shot results The Zero-shot experiments demonstrate a major improvement over the baselines, as shown in Table 5. The performance increase on the textually richer Wiki-ZS data set is considerable, and lesser on the textually sparser NELL data set. Where textual descriptions are available, both zero-shot models make significantly better predictions.

7. Discussion

In our results, we see an improvement of performance on the data sets for which we have textual descriptions. On NELL, where textual descriptions are not available for the entities, the improvement is smaller since the entities and relations are represented using much less information. The improvement is also smaller for the few-shot setting than in the zero-shot setting. The ablation study confirms this result (Table 7), where we see a clear drop in performance when we remove the textual description, in particular the label. However, we still witness a solid improvement in the zero-shot scenario, indicating that using textual features in these scenarios is highly beneficial.

We also see that for all metrics, the transformer-based embeddings provide a better result than the SIF embeddings. This may be due to a better model of phrase structure in the description texts, or in better coverage of the background text data the model's word vectors are trained on. The effect of training data coverage needs still to be determined. It is worth noting that the processing cost of using transformer models may be prohibitive, as indeed was found by us in Few-shot condition, and that models such as the more economical SIF have their place in practical use. In only one case do we see that DistMult embeddings are slightly better than the BERT embeddings in the zero-shot scenario (see Table 5).

Data set	Model	Embed method	Hits@10	Hits@5	Hits@1	MRR
NELL-ZS	OntoZSL	TransE	.352	.295	.156	.225
		DistMult	.371	.313	**.188**	.252
		SIF (Ours)	.407	.336	.184	.261
		BERT (Ours)	**.411**	**.349**	.185	**.266**
	ZS-GAN	TransE	.340	.274	.144	.211
		DistMult	.378	.310	.179	.247
		SIF (Ours)	.383	.314	.151	.232
		BERT (Ours)	**.426**	**.352**	**.200**	**.277**
Wiki-ZS	OntoZSL	TransE	.259	.211	.140	.184
		DistMult	.289	.238	.167	.211
		SIF (Ours)	.501	.411	.240	.329
		BERT (Ours)	**.604**	**.528**	**.349**	**.437**
	ZS-GAN	TransE	.245	.195	.109	.159
		DistMult	.277	.235	.158	.204
		SIF (Ours)	.475	.392	.237	.318
		BERT (Ours)	**.567**	**.489**	**.327**	**.409**

Table 5: Results on Zero-shot learning data sets. Bold marks the best results for each neural model; underlined the best for each data set.

Data set	Embeddings	Hits@10	Hits@5	Hits@1	MRR
NELL-One	TransE	.413	.343	.221	.282
	DistMult	.404	.348	.25	.301
	SIF	**.446**	**.384**	**.272**	**.328**
Wiki-One	TransE	.443	.373	.253	.313
	DistMult	.359	.303	.206	.256
	SIF	**.491**	**.422**	**.294**	**.357**

Table 6: Results from running 5-shot completion with the FAAN architecture on NELL-One and Wiki-One.

	Features	Hits@10	Hits@5	Hits@1	MRR
NELL-One	All	.437	.365	.22	.291
	No relation description	.446	.384	.272	.328
	No Category	.404	.329	.204	.268
	No Label	.303	.240	.151	.204
Wiki-One	All	.492	.422	.304	.363
	No Label	.443	.368	.242	.305
	No Description	.463	.398	.265	.323
	No Instance	.494	.429	.303	.365

Table 7: Results from the few-shot ablation studies. Removing labels and descriptions decreases the performance the most.

One limitation with our method is of course the requirement that all entities and relations have some form of textual description. Even for NELL, a data set with sparse textual descriptions, we do find a performance improvement, leading us to conclude that using text is helpful even in a low-resource setting.

Another important point to discuss is that there is a strong link between Wikidata, Wikipedia, and pre-trained models in general: as many pre-trained models are trained on the English part of Wikipedia (among other sources of text) there is a higher probability that the words for a head and a tail entity have been seen together during the training of the word vectors. While the GloVe vectors used in this were trained on a Wikipedia dump, the `Paraphrase-MiniLM-L12-v2` was trained on SimpleWiki, a version of Wikipedia with simple English. This, we argue, is precisely the advantage that textual background resources are intended to convey to our approach. The specific relations of the Knowledge Graph are *not* explicitly encoded in the trained representations, and the fact that the textual information links the head and the tail entities is what general purpose learning is all about. This motivates every textual transfer learning approach where the objective is to move from unstructured text data to a structured format, for example, in the format of a Knowledge Graph.

8. Conclusion

In this paper, we have shown how descriptive text in Knowledge Graphs is useful for inferring missing relations. Our textually enriched representation of entities and relations yields better results on two well-established benchmarks for the Few- and Zero-shot Knowledge Graph completion task. We also show that when the textual material is sparse, as in one of the benchmarks, performance improvement is lesser, although still considerable in most conditions. Our experiments bridge the gap between using structural and textual features as input to Few- and Zero-shot Knowledge Graph Completion methods and make a new enriched Knowledge Graph resource available for future research. From a more general perspective, our experiments demonstrate the value of using textual resources to enrich more formal representations of human knowledge and in the utility of transfer learning from textual data and text collections to enrich and maintain knowledge resources in general.

Acknowledgements

This work was partially supported by the Wallenberg AI Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation. Many thanks also go to Martin Isaksson, Fredrik Olsson, Tobias Norlund, and Sofie Verrewaere for valuable input and advice.

Bibliographical References

Allen, C., Balazevic, I., and Hospedales, T. (2021). Interpreting knowledge graph relation representation from word embeddings. In *International Conference on Learning Representations*.

Arora, S., Liang, Y., and Ma, T. (2019). A simple but tough-to-beat baseline for sentence embeddings. In *5th International Conference on Learning Representations (ICLR)*.

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In *Neural Information Processing Systems (NIPS)*, pages 1–9.

Chen, M., Zhang, W., Zhang, W., Chen, Q., and Chen, H. (2019). Meta relational learning for few-shot link prediction in knowledge graphs. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4217–4226, Hong Kong, China, November. Association for Computational Linguistics.

Chen, J., Geng, Y., Chen, Z., Horrocks, I., Pan, J. Z., and Chen, H. (2021a). Knowledge-aware zero-shot learning: Survey and perspective. *IJCAI 2021, Survey track*.

Chen, J., Hu, P., Jimenez-Ruiz, E., Holter, O. M., Antonyrajah, D., and Horrocks, I. (2021b). Owl2vec*: Embedding of owl ontologies. *Machine Learning*, 110(7):1813–1845.

Dettmers, T., Minervini, P., Stenetorp, P., and Riedel, S. (2018). Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June. Association for Computational Linguistics.

Finn, C., Abbeel, P., and Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*, pages 1126–1135. PMLR.

Frome, A., Corrado, G., Shlens, J., Bengio, S., Dean, J., Ranzato, M., and Mikolov, T. (2013). Devise: A deep visual-semantic embedding model.

Geng, Y., Chen, J., Chen, Z., Pan, J. Z., Ye, Z., Yuan, Z., Jia, Y., and Chen, H. (2021). Ontozsl: Ontology-enhanced zero-shot learning. In *Proceedings of the Web Conference 2021*, pages 3325–3336.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.

Hu, L., Zhang, M., Li, S., Shi, J., Shi, C., Yang, C., and Liu, Z. (2021). Text-graph enhanced knowledge graph representation learning. *Frontiers in Artificial Intelligence*, 4.

Huisman, M., van Rijn, J. N., and Plaat, A. (2021). A survey of deep meta-learning. *Artificial Intelligence Review*, pages 1–59.

Ji, S., Pan, S., Cambria, E., Marttinen, P., and Philip, S. Y. (2021). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*.

Jones, K. S. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*.

Junczys-Dowmunt, M., Grundkiewicz, R., Dwojak, T., Hoang, H., Heafield, K., Neckermann, T., Seide, F., Germann, U., Fikri Aji, A., Bogoychev, N., Martins, A. F. T., and Birch, A. (2018). Marian: Fast neural machine translation in C++. In *Proceedings of ACL 2018, System Demonstrations*, pages 116–121, Melbourne, Australia, July. Association for Computational Linguistics.

Li, J., Wang, R., Zhang, N., Zhang, W., Yang, F., and Chen, H. (2020). Logic-guided semantic representation learning for zero-shot relation classification. In *COLING*.

Lv, X., Gu, Y., Han, X., Hou, L., Li, J., and Liu, Z. (2019). Adapting meta knowledge graph information for multi-hop reasoning over few-shot relations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3376–3381, Hong Kong, China, November. Association for Computational Linguistics.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, et al., editors, *Advances in Neural Information Processing Systems*, volume 26. Curran Associates, Inc.

Min, B., Grishman, R., Wan, L., Wang, C., and Gondek, D. (2013). Distant supervision for relation extraction with an incomplete knowledge base. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 777–782.

Norouzi, M., Mikolov, T., Bengio, S., Singer, Y., Shlens, J., Frome, A., Corrado, G. S., and Dean, J. (2013). Zero-shot learning by convex combination of semantic embeddings. *arXiv preprint arXiv:1312.5650*.

Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*,

pages 1532–1543. Association for Computational Linguistics.

Qin, P., Wang, X., Chen, W., Zhang, C., Xu, W., and Wang, W. Y. (2020). Generative adversarial zero-shot relational learning for knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8673–8680.

Reimers, N. and Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 11.

Ruffinelli, D., Broscheit, S., and Gemulla, R. (2019). You can teach an old dog new tricks! on training knowledge graph embeddings. In *International Conference on Learning Representations*.

Sheng, J., Guo, S., Chen, Z., Yue, J., Wang, L., Liu, T., and Xu, H. (2020). Adaptive Attentional Network for Few-Shot Knowledge Graph Completion. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1681–1691, Online, November. Association for Computational Linguistics.

Snell, J., Swersky, K., and Zemel, R. S. (2017). Prototypical networks for few-shot learning. *arXiv preprint arXiv:1703.05175*.

Socher, R., Chen, D., Manning, C. D., and Ng, A. (2013). Reasoning with neural tensor networks for knowledge base completion. In *Advances in neural information processing systems*, pages 926–934.

Sun, Z., Deng, Z.-H., Nie, J.-Y., and Tang, J. (2019). Rotate: Knowledge graph embedding by relational rotation in complex space. *CoRR*, abs/1902.10197.

Vinyals, O., Blundell, C., Lillicrap, T., kavukcuoglu, k., and Wierstra, D. (2016). Matching networks for one shot learning. In D. Lee, et al., editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

Wang, X., Gao, T., Zhu, Z., Zhang, Z., Liu, Z., Li, J., and Tang, J. (2021). Kepler: A unified model for knowledge embedding and pre-trained language representation. *Transactions of the Association for Computational Linguistics*, 9:176–194.

Xiao, H., Huang, M., Meng, L., and Zhu, X. (2017). SSP: Semantic space projection for knowledge graph embedding with text descriptions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Xie, R., Liu, Z., Jia, J., Luan, H., and Sun, M. (2016). Representation learning of knowledge graphs with entity descriptions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.

Xiong, W., Yu, M., Chang, S., Guo, X., and Wang, W. Y. (2018). One-shot relational learning for knowledge graphs. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1980–1990, Brussels, Belgium, October-November. Association for Computational Linguistics.

Yang, B., tau Yih, W., He, X., Gao, J., and Deng, L. (2015). Embedding entities and relations for learning and inference in knowledge bases. *CoRR*, abs/1412.6575.

Zhang, C., Yao, H., Huang, C., Jiang, M., Li, Z., and Chawla, N. V. (2020). Few-shot knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3041–3048.

Lexicographical References

Cornell, F. (2022). Challenging the assumption of structure-based embeddings in few- and zero-shot knowledge graph completion. `https://github.com/Filco306/challenging-structural-assumptions/`.

Mitchell, T., Cohen, W., Hruschka, E., Talukdar, P., Yang, B., Betteridge, J., Carlson, A., Dalvi, B., Gardner, M., Kisiel, B., Krishnamurthy, J., Lao, N., Mazaitis, K., Mohamed, T., Nakashole, N., Platanios, E., Ritter, A., Samadi, M., Settles, B., Wang, R., Wijaya, D., Gupta, A., Chen, X., Saparov, A., Greaves, M., and Welling, J. (2018). Never-ending learning. *Commun. ACM*, 61(5):103–115, April.

Vrandečić, D. and Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. *Communications of the ACM*, 57(10):78–85.