Knowledge Graph Embedding with Numeric Attributes of Entities

Yanrong Wu, Zhichun Wang*
College of Information Science and Technology
Beijing Normal University, Beijing 100875, PR. China
yrwu@mail.bnu.edu.cn, zcwang@bnu.edu.cn

Abstract

Knowledge Graph (KG) embedding projects entities and relations into low dimensional vector space, which has been successfully applied in KG completion task. The previous embedding approaches only model entities and their relations, ignoring a large number of entities' numeric attributes in KGs. In this paper, we propose a new KG embedding model which jointly model entity relations and numeric attributes. Our approach combines an attribute embedding model with a translation-based structure embedding model, which learns the embeddings of entities, relations, and attributes simultaneously. Experiments of link prediction on YAGO and Freebase show that the performance is effectively improved by adding entities' numeric attributes in the embedding model.

1 Introduction

Recently, a number of Knowledge Graphs (KGs) have been created, such as DBpedia (Lehmann, 2015), YAGO (Mahdisoltani et al., 2015), and Freebase (Bollacker et al., 2008). KGs encode structured information of entities in the form of triplets (e.g. $\langle Microsoft, isLocatedIn, UnitedStates\rangle$), and have been successfully applied in many real-world applications. Although KGs contain a huge amount of triplets, most of them are incomplete. In order to further expand KGs, much work on KG completion has been done, which aims to predict new triplets based on the existing ones in KGs. A promising group of research for KG completion is known as KG embedding. KG embedding approaches project entities and relations into a continuous vector space while preserving the original knowledge in the KG. KG embedding models achieve good performance in KG completion in terms of efficiency and scalability. TransE is a representative KG embedding approach (Bordes et al., 2013), which projects both entities and relations into the same vector space: if a triplet $\langle head\ entity, relation, tail\ entity\rangle$ (denoted as $\langle h, r, t\rangle$) holds, TransE wants that $\boldsymbol{h} + \boldsymbol{r} \approx \boldsymbol{t}$. The embeddings are learned by minimizing a margin-based ranking criterion over the training set. TransE model is simple but powerful, and it gets promising results on link prediction and triple classification problems. There are several enhanced model of TransE, including TransR (Lin et al., 2015), TransH (Wang et al., 2014) and TransD (Ji et al., 2015) etc. By introducing new representations of relational translation, later approaches achieve better performance at the cost of increasing model complexity. Recent surveys (Wang et al., 2017; Nickel et al., 2016) give detailed introduction and comparison of various KG embedding approaches.

However, most of the existing KG embedding approaches only model relational triplets (i.e. triplets of entity relations), while ignoring a large number of attributive triplets (i.e. triplets of entity attributes, e.g. $\langle Microsoft, wasFoundedOnDate, 1975\rangle$) in KGs. attributive triplets describe various attributes of entities, such as ages of people or areas of a city. There are a huge number of attributive triplets in real KGs, and we believe that information encoded in these triplets is also useful for predicting entity relations. Having the above motivation, we propose a new KG embedding approach that jointly model entity relations and entities' numeric attributes. Our approach consists of two component models,

*Corresponding Author

Proceedings of the 3rd Workshop on Representation Learning for NLP, pages 132–136
Melbourne, Australia, July 20, 2018.

structure embedding model and attribute embedding model. The structure embedding model is a translational distance model that preserves the knowledge of entity relations; the attribute embedding model is a regression-based model that preserves the knowledge of entity attributes. Two component models are jointly optimized to get the embeddings of entities, relations, and attributes. Experiments of link prediction on YAGO and Freebase show that the performance is effectively improved by adding entities' numeric attributes in the embedding model.

2 Our Approach

To effectively utilize numeric attributes of entities in KG embedding, we propose **TransEA**, which combine a new attribute embedding model with the structure embedding model of TransE. Two component models in TransEA share the embeddings of entities, and they are jointly optimized in the training process.

2.1 Structure Embedding

The structure embedding directly adopts the translation-based method in TransE to model the relational triplets in KGs. Both Entities and relations in a KG are represented in the same vector space $\mathbb{R}^d$. In a triplet $\langle h, r, t\rangle$, the relation is considered as a translation vector $\mathbf{r}$, which connects the vector of entities $\mathbf{h}$ and $\mathbf{t}$ with low error, i.e. $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$. The score function of a given triplet $\langle h, r, t\rangle$ is defined as

$$f_r(h, t) = -||\mathbf{h} + \mathbf{r} - \mathbf{t}||_{1/2} \tag{1}$$

$||x||_{1/2}$ denotes either the $L1$ or $L2$ norm. For all the relational triplets in the KG, the loss function of the structure embedding is defined as:

$$L_R = \sum_{\langle h,r,t\rangle \in S} \sum_{\langle h',r,t'\rangle \in S'} [\gamma + f_r(h, t) - f_r(h', t')]_+ \tag{2}$$

where $[x]_+ = \max\{0, x\}$, S' denotes the set of negative triplets constructed by corrupting $\langle h, r, t\rangle$, i.e. replacing h or t with a randomly chosen entity in KG; $\gamma > 0$ is a margin hyper-parameter separating positive and negative triplets.

2.2 Attribute Embedding

Attribute embedding model takes all the attributive triplets in a KG as input, and learns embeddings of entities and attributes. Both entities and attributes are represented as vectors in space $\mathbb{R}^d$. In an attributive triplet $\langle e, a, v\rangle$, e is an entity, a is an attribute, and v is the value of the entity's attribute. In our approach, we only consider attributive triplets containing numeric values or values can be easily converted into numeric ones. For a triplet $\langle e, a, v\rangle$, we define a score function as

$$f_a(e, v) = -||\mathbf{a}^\top \cdot \mathbf{e} + b_a - v||_{1/2} \tag{3}$$

where $\mathbf{a}$ and $\mathbf{e}$ are vectors of attribute a and entity e, b_a is a bias for attribute a. The idea of this score function is to predict the attribute value by a linear regression model of attribute a; the vector $\mathbf{a}$ and bias b_a are the parameters of the regression model. For all the attributive triplets in the KG, the loss function of the attribute embedding is defined as:

$$L_A = \sum_{\langle e,a,v\rangle \in T} f_a(e, v) \tag{4}$$

where T is the set of all attributive triplets with numeric values in the KG.

2.3 Joint Model

To combine the above two component models, TransEA minimizes the following loss function:

$$L = (1 - \alpha) \cdot L_R + \alpha \cdot L_A \tag{5}$$

where α is a hyper-parameter that balances the importance of structure and attribute embedding. In the joint model, we let the embeddings of entities shared by two component models. Entities, relations, and attributes are all represented by vectors in $\mathbb{R}^d$. We implement our approach by using TensorFlow[1], and the loss function is minimized by performing stochastic gradient descent.

3 Experiments

3.1 Datasets

The following two datasets are used in the experiments, Table 1 shows their detail information.

YG58K. YG58K is a subset of YAGO3 (Mahdisoltani et al., 2015) which contains about 58K entities. YG58K is built by removing entities from YAGO3 that appear less than 25 times or have no attributive triplets. All the remaining triplets are then randomly split into training/validation/test sets.

[1] https://www.tensorflow.org

FB15K. FB15K is a subset of triplets extracted from Freebase[2]. This subset of Freebase was originally used in (Bordes et al., 2013), and then widely used for evaluating KB completion approaches. Since our approach consumes attributive triplets, we extract all the attributive triplets of entities in FB15K from Freebase to build the evaluation dataset.

Datasets	YG58K	FB15K
# Relational Triplets	497783	592213
# Attributive Triplets	130287	24034
# Entities	58130	14951
# Relations	32	1345
# Attributes	24	336
# Train Sets	399480	483142
# Valid Sets	49171	59071
# Test Sets	49132	50000

Table 1: Statistics of datasets

3.2 Experimental setup

In the experiments, Mean Rank (the mean rank of the original correct entity), Hits@k (the proportion of the original correct entity to the top k entities), and MRR (the mean reciprocal rank) are used as evaluation metrics. Given a testing triplet $\langle h, r, t\rangle$, we replace the head h by every entity in the KGs and calculate dissimilarity measures according to the score function f_r. Ranking the scores in ascending order, then we get the rank of the original correct triplet to compute the evaluation metrics. And we repeat the procedure when removing the tail t instead of the head h. We name the evaluation setting as **"Raw"**. While corrupted triplets that appear in the train/valid/test sets (except the original correct one) may underestimate the metrics, we also filter out those corrupted triplets before getting the rank of each testing triplet and we call this process **"Filter"**.

Because our approach is built based on TransE, we compare our approach with TransE to see whether adding attribute embedding in the model improves the performance of link prediction. For TransE and TransEA, we consider the learning rate λ among {0.1, 0.01, 0.001}, the margin γ among {1, 2, 4, 10}, the dimensions of embedding d among {20, 50, 100, 150}, the types of norm in two score functions among {L1, L2}, and α among {0.2, 0.3, 0.4, 0.5, 0.6}. Based on the mean rank in validation set, we select the best configurations for two approaches. On the YG58K dataset, the best parameter configuration for TransE is $(\lambda = 0.1, \gamma = 4, d = 50, f_r = L1, f_a = L1)$, and for TransEA is $(\lambda = 0.001, \gamma = 4, d = 50, f_r = L1, f_a = L1, \alpha = 0.6)$. On the FB15K dataset, the best parameter configuration for TransE is$(\lambda = 0.01, \gamma = 1, d = 50, f_r = L1, f_a = L1)$, and for TransEA is $(\lambda = 0.001, \gamma = 2, d = 100, f_r = L1, f_a = L1, \alpha = 0.3)$.

[2]https://everest.hds.utc.fr/doku.php?id=en:transe

3.3 Results

Table 2 shows the results of link prediction on YG58K and FB15K datasets. The results of predicting head and tail entities are outlined separately, and we also report the overall results by considering prediction of both head and tail entity. According to the overall results, TransEA outperforms TransE on both two datasets in terms of all the three metrics. TransEA gets lower Mean Ranks by about 10 on YG58K dataset; the MRR and Hits@k of two approaches are very close, TransEA gets slightly better results, the improvements of MRR and Hits@k are 0.1-0.2% and 0-0.3%. On FB15K dataset, TransEA gets lower Mean Ranks by 13, and it also gets better results than TransE according to MRR, Hits@10 and Hits@3.

Table 3 shows the results of different relational categories. In general, TransEA has superiority on two datasets, except one-to-many relation for replacing head entity on YG58K. And the improvements on FB15K are larger than YG58K.

In order to figure out which relations are predicted more accurately by TransEA, Table 4 lists the top 5 improved relations in terms of Hits@10 on YG58K. It shows the best improvement of Hits@10 is 25% for the relation `isInterestedIn`. The second one is 12.5% for `hasAcademicAdvisor`, and the third is 6.3% for `worteMusicFor`. Entities of these three relations have plenty of numeric attributes (`wasBornOnDate`, `diedOnDate`) describing people, we believe they are helpful to improving the embeddings of entity relations. Entities in relational triplets about `livesIn`, (e.g. $\langle HankAzaria, livesIn, NewYork\rangle$), also have some numeric attributes (`hasLatitude`, `hasLongtude`, `hasNumberOfPeople`, `etc`), therefore TransEA gets a 5% improvement of Hits@10.

On FB15K dataset, five relations have 100%

Dataset	Entity	Model	Mean Rank		MRR(%)		Hits@10(%)		Hits@3(%)		Hits@1(%)	
			Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter	Raw	Filter
YG58K	Head	TransE	950	731	3.1	5.2	9.1	15.4	**4.1**	8.4	1.0	3.2
		TransEA	**944**	**723**	**3.1**	**5.4**	**9.4**	**16.0**	4.1	**8.5**	**1.1**	**3.4**
	Tail	TransE	240	234	8.4	10.2	27.0	31.9	12.2	17.0	4.5	6.5
		TransEA	**229**	**223**	**8.5**	**10.5**	**27.6**	**32.7**	**12.4**	**17.6**	**4.7**	**6.8**
	All	TransE	595	482	5.7	7.7	18.0	23.7	8.2	12.7	2.8	4.8
		TransEA	**586**	**473**	**5.8**	**7.9**	**18.5**	**24.3**	**8.2**	**13.0**	**2.9**	**5.1**
FB15K	Head	TransE	240	115	14.5	25.2	47.0	68.7	26.2	52.4	11.8	30.6
		TransEA	**225**	**100**	**15.1**	**28.1**	**49.5**	**74.0**	**28.0**	**60.1**	**11.8**	**34.8**
	Tail	TransE	168	87	17.6	28.2	54.8	75.1	32.8	58.9	**16.4**	35.7
		TransEA	**157**	**76**	**18.2**	**30.9**	**57.5**	**80.5**	**34.5**	**66.6**	16.3	**40.0**
	All	TransE	204	101	16.0	26.7	50.9	71.9	29.5	55.7	**14.1**	33.1
		TransEA	**191**	**88**	**16.7**	**29.5**	**53.5**	**77.3**	**31.3**	**63.3**	14.0	**37.4**

Table 2: Link prediction results

DATASETS	TASK	Predicting Head(Hits@10)				Prediction Tail(Hits@10)			
	REL.CAT	1-to-1	1-to-N	N-to-1	N-to-N	1-to-1	1-to-N	N-to-1	N-to-N
YG58K	TransE	61.4	**45.5**	15.4	15.5	62.0	18.2	31.9	31.1
	TransEA	**63.9**	36.4	**16.0**	**16.0**	**63.3**	**22.7**	**32.7**	**31.9**
FB15K	TransE	78.1	93.8	68.7	72.3	78.0	42.1	75.1	75.6
	TransEA	**84.3**	**95.5**	**74.0**	**77.6**	**83.3**	**52.4**	**80.5**	**81.1**

Table 3: Hits@10(%) by relational category in the filtered evaluation setting. (N. stand for MANY)

Relation	TransE	TransEA
isInterestedIn	50.0	75.0
hasAcademicAdvisor	31.3	43.8
wroteMusicFor	12.5	18.8
livesIn	23.8	28.8
hasNeighbor	48.1	52.8

Table 4: Top 5 relations of promoted Hits@10 and their Hits@10(%) on YG58K

Relation	TransE	TransEA
business/brand/company	24	2
base/celebrity/restaurant	249	4
base/celebrity/product	24	2
music/artists_supported	44	3
sports/competition/country	24	4

Table 5: Top 5 relations of promoted Hit@10 and their Mean Rank on FB15K

improvements of Hits@10, because TransE does not correctly predict any correct triplets in the top 10 ranked ones. We find that these relations only have one single sample in the test sets, so Table 5 lists the Mean Rank of them. Obviously, TransEA improves their Mean Rank a lot. Entities in triplets of the five relations have only a few attributes. For example, the relation `business/brand/company` only has one numeric attributive triplet about `organization/dateFounded`. And the relation `music/artists supported` has two triplets with numeric attributes `person/dateOfBirth` and one triplet with `person/heightMeters`. Therefore, the quality of predicted links can be improved as well even with only a small number of entities numeric attributes.

4 Conclusion

In this paper, we propose TransEA, an embedding approach which jointly models relational and attributive triplets in KGs. TransEA combines an attribute embedding model with the translation-based embedding model in TransE. Experiments on YAGO and Freebase show that TransEA achieves better performance than TransE in link prediction task. In the future, we will study how to predict missing attribute values in KGs based on KG embedding.

Acknowledgments

The work is supported by the National Key Research and Development Program of China (No. 2017YFC0804004) and the National Natural Science Foundation of China (No. 61772079).

References

Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. *SIGMOD 08 Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250.

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Proceedings of Advances in neural information processing systems (NIPS2013)*, pages 2787–2795.

Guoliang Ji, Shizhu He, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Knowledge graph embedding via dynamic mapping matrix. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, volume 1, pages 687–696.

J. Lehmann. 2015. Dbpedia: A large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web*, 6(2):167–195.

Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI2015)*, volume 15, pages 2181–2187.

Farzaneh Mahdisoltani, Joanna Asia Biega, and Fabian M. Suchanek. 2015. YAGO3: A Knowledge Base from Multilingual Wikipedias . *7th Biennial Conference on Innovative Data Systems Research*.

Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. 2016. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33.

Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. 2017. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743.

Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. 2014. Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the Twenty-eighth AAAI Conference on Artificial Intelligence (AAAI2014)*, volume 14, pages 1112–1119.