



Composing Relationships with Translations

Alberto García-Durán, Antoine Bordes, Nicolas Usunier

► To cite this version:

Alberto García-Durán, Antoine Bordes, Nicolas Usunier. Composing Relationships with Translations. Conference on Empirical Methods in Natural Language Processing (EMNLP 2015), Sep 2015, Lisbonne, Portugal. pp.286-290, 10.18653/v1/D15-1034 . hal-01301243

HAL Id: hal-01301243

<https://hal.science/hal-01301243>

Submitted on 11 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Composing Relationships with Translations

Alberto García-Durán
Sorbonne universités, UTC
CNRS, Heudiasyc 7253
60203 Compiègne, France
agarciad@utc.fr

Antoine Bordes
Facebook AI Research
770 Broadway
New York, NY, USA
abordes@fb.com

Nicolas Usunier
Facebook AI Research
112, avenue de Wagram
75017 Paris, France
usunier@fb.com

Abstract

Performing link prediction in Knowledge Bases (KBs) with embedding-based models, like with the model TransE (Bordes et al., 2013) which represents relationships as translations in the embedding space, have shown promising results in recent years. Most of these works focused on modeling single relationships and hence do not take full advantage of the graph structure of KBs. In this paper, we propose an extension of TransE that learns to explicitly model composition of relationships via the addition of their corresponding translation vectors. We show empirically that this allows to improve performance for predicting single relationships as well as compositions of pairs of them.

1 Introduction

Performing link prediction on multi-relational data is becoming essential in order to complete the huge amount of missing information of the knowledge bases. These knowledge can be formalized as directed multi-relation graphs, whose node correspond to entities connected with edges encoding various kind of relationships. We denote these connections via triples (*head*, *label*, *tail*). Link prediction consists in filling in incomplete triples like (*head*, *label*, *?*) or (*?*, *label*, *tail*).

In this context, embedding models (Wang et al., 2014; Lin et al., 2015b; Jenatton et al., 2012; Socher et al., 2013) that attempts to learn low-dimensional vector or matrix representations of entities and relationships have shown promising performance in recent years. In particular, the basic model TransE (Bordes et al., 2013) has been proved to very powerful. This model treats each relationship as a translation vector operating on the embedding representing the entities. Hence,

for a triple (*head*, *label*, *tail*), the vector embeddings of *head* and *tail* are learned so that they are connected through a translation parameterized by the vector associated with *label*. Many extensions have been proposed to improve the representation power of TransE while still keeping its simplicity, by adding some projections steps before the translation (Wang et al., 2014; Lin et al., 2015b).

In this paper, we proposed an extension of TransE¹ that focuses on improving its representation of the underlying graph of multi-relational data by trying to learn compositions of relationships as sequences of translations in the embedding space. The idea is to train the embeddings by learning simple reasonings, such as the relationship *people/nationality* should give a similar result as the composition *people/city_of_birth* and *city/country*. In our approach, called RTransE, the training set is augmented with relevant examples of such compositions, and training so that sequences of translations lead to the desired result. This idea of compositionality with translations has been simultaneously and in complete independence developed in (Lin et al., 2015a). The main difference between both works relies in the way of handle composition: whereas they do it through a ranking criterion, we do it through regularization.

We present experimental results that show the superiority of RTransE over TransE in terms of link prediction. A detailed evaluation, in which test examples are classified as *easy* or *hard* depending on their similarity with training data, highlights the accuracy of RTransE on the harder examples. Our experiments include a new evaluation protocol, in which the model is directly asked to answer questions related to compositions of relations, such as (*head*, *label*₁, *label*₂, *?*). RTransE also achieves significantly better performances than TransE on this new dataset.

¹Code available in <https://github.com/glorotxa/SME>

We describe RTRANSE in the next section, and present our experiments in Section 3.

2 Model

The model we propose is inspired by TRANSE (Bordes et al., 2013). In TRANSE, entities and relationships of a KB are mapped to low dimensional vectors, called embeddings. These embeddings are learnt so that for each fact (h, ℓ, t) in the KB, we have $h + \ell \approx t$ in the embedding space.

Using translations for relationships naturally leads to embed the composition of two relationships as the sum of their embeddings: on a path $(h, \ell, t), (t, \ell', t')$, we should have $h + \ell + \ell' \approx t'$ in the embedding space. The original TRANSE does not enforce that the embeddings accurately reproduce such compositions. The *recurrent* TRANSE we propose here has a modified training stage to include such compositions. This should allow to model simple reasonings in the KB, such as *people/nationality* is similar to the composition of *people/city_of_birth* and *city/country*.

2.1 Recurrent TransE

We describe in this section our model in its full generality, which allows to deal with compositions of an arbitrary number of relationships, even though in this first work we experimented only with compositions of two relationships.

Triples that are the result of a compositions are denoted by $(h, \{\ell_i\}_{i=1}^p, t)$, where p is the number of relationships that are composed to go from h to t . Such a path means that there exist entities e_1, \dots, e_{p+1} , with $e_1 = h$ and $e_{p+1} = t$ such that for all k , (e_k, ℓ_k, e_{k+1}) is a fact in the KB. Our model, RTRANSE for *recurrent* TRANSE, represents each step $s_k(h, \{\ell_i\}_{i=1}^p, t)$ along the path in the KB with the recurrence relationship (boldface characters denote embedding vectors):

$$\begin{aligned} s_1(h, \{\ell_i\}_{i=1}^p, t) &= \mathbf{h} \\ s_{k+1}(h, \{\ell_i\}_{i=1}^p, t) &= s_k(h, \{\ell_i\}_{i=1}^p, t) + \ell_k. \end{aligned}$$

Then, the energy of a triple is computed as

$$d(h, \{\ell_i\}_{i=1}^p, t) = \|\mathbf{s}_p(h, \{\ell_i\}_{i=1}^p, t) - \mathbf{t}\|_2.$$

2.2 Path construction and filtering

The experience of the paper is motivated by learning simple reasonings in the KB through the compositions of relationships. Therefore, we restricted our analysis to paths of length 2 created as follows.

First, for each fact (h, ℓ, t) , retrieve all paths $(h, \{\ell_1, \ell_2\}, t)$ such that there is e such that both (h, ℓ_1, e) and (e, ℓ_2, t) are in the KB. Then, we filter out paths where $(h, \ell_1, e) = (h, \ell, t)$ or $(e, \ell_2, t) = (h, \ell, t)$, as well as the paths with $\ell_1 = \ell_2$ and $h = e = t$.

We focused on “unambiguous” paths, so that the reasoning might actually make sense. In particular, we considered only paths where ℓ_1 is either a 1-to-1 or a 1-to-many relationship, and where ℓ_2 is either a 1-to-1 or a many-to-1 relationship. In our experiments, the paths created for training only consider the training subset of facts.

In the remainder of the paper, such paths of length 2 are called *quadruples*.

2.3 Training and regularization

Our training objective is decomposed in two parts: the first one is the ranking criterion on triples of TRANSE, ignoring quadruples. Paths are then taken into account through additional regularization terms.

Denoting by \mathcal{S} the set of facts in the KB, the first part of the training objective is the following ranking criterion that operates on triples

$$\sum_{\substack{(h, \ell, t) \in \mathcal{S} \\ (h', \ell', t') \in \mathcal{S}_{(h, \ell, t)}}} [\gamma + d(h, \ell, t) - d(h', \ell', t')]_+,$$

where $[x]_+ = \max(x, 0)$ is the positive part of x , γ is a margin hyperparameter and $\mathcal{S}_{(h, \ell, t)}$ is the set of corrupted triples created from (h, ℓ, t) by replacing either h or t with another KB entity.

This ranking loss effectively trains so that the embedding of the tail is the nearest neighbor of the translated head, but it does not guarantee that the distance between the tail and the translated head is small. The nearest neighbor criterion is sufficient to make inference over simple triples, but making sure that the distance is small is necessary for the composition rule to be accurate. In order to account for the compositionality of relationships, we add two additional regularization terms:

- $\lambda \sum_{(h, \ell, t) \in \mathcal{S}} d(h, \ell, t)^2$
- $\alpha \sum_{(h, \{\ell_1, \ell_2\}, t) \in \mathcal{S}} N_{\ell \rightarrow \{\ell_1, \ell_2\}} d(h, \{\ell_1, \ell_2\}, t)^2.$

The first criterion only applies to original facts of the KB, while the second term applies to quadruples. $N_{\ell \rightarrow \{\ell_1, \ell_2\}}$, which involves both the

DATA SET	FAMILY	FB15K
ENTITIES	721	14,951
RELATIONSHIPS	7	1,345
TRAINING TRIPLES	8,461	483,142
TRAINING QUAD.	–	30,252
VALIDATION TRIPLES	2,820	50,000
TEST TRIPLES	2,821	59,071
TEST QUAD.	–	1,852

Table 1: Statistics of the datasets.

MODEL	TRANSE		RTRANSE	
	MR	H@10	MR	H@10
EASY	17.7	76.8	12.5	82.2
HARD	191.0	48.9	205.7	51.0
EASY W. COMP.	16.4	78.8	11.6	83.0
EASY W/O COMP.	21.6	71.3	16.0	75.3
HARD W. COMP.	208.1	46.8	212.2	49.3
HARD W/O COMP.	122.9	57.0	123.8	57.5
OVERALL	50.7	71.5	49.5	76.2

Table 2: **Detailed performances on FB15k** of TRANSE and RTRANSE. H@10 are in %. W. COMP. indicates examples for which there exist quadruplets in train matching their relationship.

relationships of the quadruple and the relationship ℓ from which it was created, is the number of paths involving relationships $\{\ell_1, \ell_2\}$ created from a fact involving ℓ , normalized by the total number of quadruples created from facts involving ℓ . This criterion puts more weight on paths that are reliable as an alternative for a relationship, for instance $\{people/city_of_birth, city/country\}$ is likely a better alternative to $people/nationality$ than $\{people/writer_of_the_film, film/film_release_region\}$. Finally, a regularization term $\mu||e||_2^2$ is added for each entity embedding e .

3 Experiments

This section presents experiments on the benchmark FB15K introduced in (Bordes et al., 2013) and on FAMILY, a slightly extended version of the artificial database described in (García-Durán et al., 2014). Table 1 gives their statistics.

3.1 Experimental Protocol

Data FB15K is a subset of Freebase, a very large database of generic facts gathering more than 1.2 billion triples and 80 million entities. Inspired by (Hinton, 1986), FAMILY is a database that contains triples expressing family relationships (cousin_of, has_ancestor, married_to, parent_of, related_to, sibling_of, uncle_of) among the members of 5 families along 6 generations. This dataset is artificial and each family is organized in a layered tree structure where each layer refers to a gen-

eration. Families are connected among them by marriage links between two members, randomly sampled from the same layer of different families. Interestingly on this dataset, there are obvious compositional relationships like *uncle_of* \approx *sibling_of* + *parent_of* or *parent_of* \approx *married_to* + *parent_of*, among others.

Setting Our main comparison is TRANSE so we followed the same experimental setting as in (Bordes et al., 2013), using ranking metrics for evaluation. For each test triple we replaced the head by each of the entities in turn, and then computed the score of each of these candidates and sorted them. Since other positive candidates (i.e. entities forming true triples) can be ranked higher than the target one, we filtered out all the positive candidates existing in either the training, validation and test set, except the target one, from the ranking and then we kept the rank of the target entity. The same procedure is repeated but removing the tail instead of the head. The filtered mean rank (mean rank in the rest) is the average of these ranks, and the filtered Hits@10 (H@10 in the rest) is the proportion of target entities in the top 10 predictions.

The embedding dimensions were set to 20 for FAMILY and 100 for FB15K. Training was performed by stochastic gradient descent, stopping after for 500 epochs. On FB15K, we used the embeddings of TRANSE to initialize RTRANSE, and we set a learning rate of 0.001 to fine-tune RTRANSE. On FAMILY, both algorithms were initialized randomly and used a learning rate of 0.01. The mean rank was used as a validation criterion, and the values of γ , λ , α and μ were chosen respectively among $\{0.25, 0.5, 1\}$, $\{1e^{-4}, 1e^{-5}, 0\}$, $\{0.1, 0.05, 0.1, 0.01, 0.005\}$ and $\{1e^{-4}, 1e^{-5}, 0\}$.

3.2 Results

Overall performances Experiments on FAMILY show a quantitative improvement of the performance of RTRANSE : where TRANSE gets a mean rank of 6.7 and a H@5 of 68.7, RTRANSE get a performance of 6.3 and 72.3 respectively.

Similarly, on FB15K, Table 2 (last row) shows that training on longer paths (length 2 here) actually consistently improves the performance while predicting heads and tails of triples only: the overall H@10 improves by almost 5% from 71.5 for TRANSE to 76.2 for RTRANSE.

Detailed results In order to better understand the gains of RTRANSE, we performed a detailed

	3 Nearest entities to $h + l_1 + l_2$	
	RTRANSE	TRANSE
h: madtv l_1 : regular TV app. l_2 : nationality	U.S.A. Ireland Japan	Ireland U.S.A. U.K.
h: stargate atlantis l_1 : regular TV app. l_2 : nationality	Hawaii Scotland U.S.A.	Scotland Hawaii U.K.
h: law&order l_1 : spin off l_2 : program creator	dick wolf walon green ken burns	michael crichton renny harlin paul schrader

Table 3: **Examples of predictions on quadruples** of TRANSE and RTRANSE

evaluation on FB15K, by classifying the test triples along two axes: *easy* vs *hard* and *with composition* vs *without composition*. A test triple (h, ℓ, t) is *easy* if its head and tail are connected by a triple in the training set, i.e. if either (h, ℓ', t) or (t, ℓ', h) is seen in train for some relationship ℓ' . Otherwise, the triple is *hard*. Orthogonally, the test triple (h, ℓ, t) is *with composition* if there is a path $(h, \{\ell_1, \ell_2\}, t)$ that can be constructed from the training set (notice that $(h, \{\ell_1, \ell_2\}, t)$ will usually not be used for training because training paths are built upon training triples). If no such path exists, (h, ℓ, t) is *without composition*.

The detailed results are shown in Table 2. We can see that comparatively to TRANSE, RTRANSE particularly improves performances in terms of H@10 on triples *with composition*, improving on *easy* triples by 4.2% (from 78.8% to 83.0%) and *hard* triples by 2.5% (from 46.8% to 49.3%). The main gains are still on *easy* triples, and in fact the H@10 on *easy* triples *without composition* increases by 4%, from 71.3% to 75.3%. The mean rank also considerably improves on *easy* triples, and stays somehow still on *hard* ones. All in all, the results show that considering paths during training very significantly improves performances, and the results on triples *with composition* suggest that RTRANSE is indeed capable of capturing the evidence of links that exist in longer paths.

3.3 Results on quadruples

While usual evaluations for link prediction in KBs focus on predicting a missing element of a test triple, we propose here to extend the evaluation to answering more complex questions, such as $(h, \{\ell_1, \ell_2\}, ?)$ or $(?, \{\ell_1, \ell_2\}, t)$.

Examples Table 3 presents examples of predictions of both TRANSE and our model RTRANSE on such quadruples. The two first examples try to predict the origin of a TV series from the nationality of its actors. In the first one, RTRANSE is able to correctly infer the country of origin, whereas TRANSE is affected by that fact that one of its main characters, Crista Flanagan, is American of Irish descent. In the second one, given the multiculturalism of the cast, both models list the countries of the members of the series.

Prediction performance For a more quantitative analysis, we have generated a new test dataset of link prediction on quadruples on FB15K. This test set was created by generating the paths the usual test set (the triple test set) and removing those quadruples that are used for training. We obtain 1,852 quadruples. The overall experimental protocol is the same as before, trying to predict the head or tail of these quadruple in turn.

On that evaluation protocol, RTRANSE has a mean rank of 114.0 and a H@10 of 68.2%, while TRANSE obtains a mean rank of 159.9 and a H@10 of 65.2% (using the same models as in the previous subsection). We can see that learning on paths improves performances on both metrics, with a gain of 3% in terms of H@10 and an important gain of about 46 in mean rank, which corresponds to a relative improvement of about 30%.

4 Conclusion

We proposed to learn embeddings of compositions of relationships in the translation model for link prediction in KBs. Our experimental results show that this approach is promising.

We considered in this work a restricted set of

small paths of length two. We leave the study of more general paths to future work.

Acknowledgments

This work was carried out in the framework of the Labex MS2T (ANR-11-IDEX-0004-02), and was funded by the French National Agency for Research (EVEREST-12-JS02-005-01).

References

- Antoine Bordes, Nicolas Usunier, Alberto García-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*, pages 2787–2795.
- Alberto García-Durán, Antoine Bordes, and Nicolas Usunier. 2014. Effective blending of two and three-way interactions for modeling multi-relational data. In *ECML PKDD 2014*. Springer Berlin Heidelberg.
- Geoffrey E Hinton. 1986. Learning distributed representations of concepts. In *Proceedings of the eighth annual conference of the cognitive science society*, volume 1, page 12. Amherst, MA.
- Rodolphe Jenatton, Nicolas Le Roux, Antoine Bordes, and Guillaume Obozinski. 2012. A latent factor model for highly multi-relational data. In *NIPS 25*.
- Yankai Lin, Zhiyuan Liu, and Maosong Sun. 2015a. Modeling relation paths for representation learning of knowledge bases. *arXiv preprint arXiv:1506.00379*.
- Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. 2015b. Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of AAAI’15*.
- Richard Socher, Danqi Chen, Christopher D. Manning, and Andrew Y. Ng. 2013. Reasoning With Neural Tensor Networks For Knowledge Base Completion. In *Advances in Neural Information Processing Systems 26*.
- 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*, pages 1112–1119.