Regularizing Relation Representations by First-order Implications (Abstract)

Thomas Demeester THOMAS.DEMEESTER@INTEC.UGENT.BE
IBCN, Ghent University - iMinds, Ghent, Belgium

Tim Rocktäschel, Sebastian Riedel {T.ROCKTASCHEL, S.RIEDEL}@CS.UCL.AC.UK
Machine Reading Lab, University College London, London, UK

Keywords: matrix factorization, relation extraction, first-order logic

Abstract

Methods for automated knowledge base construction often rely on trained vector representations of relations and entities. Recent work shows that such representations can be regularized to inject first-order logic formulae. This enables incorporating domain-knowledge for the improved prediction of facts. However, current approaches rely on the propositionalization of formulae and thus do not scale to large sets of formulae or knowledge bases with many facts. We propose a method that imposes first-order constraints directly on relation representations, avoiding the costly grounding of formulae. We show that our approach works well for implications between pairs of relations on artificial datasets.

1. Introduction

Many methods for automated knowledge base (KB) construction rely on learned relation and entity vector representations (Nickel et al., 2015). Such representations are hard to learn for relations with only few supporting facts in KBs. Recent work shows that the injection of additional common-sense knowledge under the form of rules is capable to improve inference on KBs (Rocktäschel et al., 2015; Wang et al., 2015). One type of rules are first-order implications, which can be manually constructed, or mined from ontologies such as WordNet (Miller, 1995). The proposed techniques are however not scalable towards large quantities of such rules, as they require the grounding of the rules in terms of the training facts.

Appearing in *Proceedings of Benelearn 2016*. Copyright 2016 by the author(s)/owner(s).

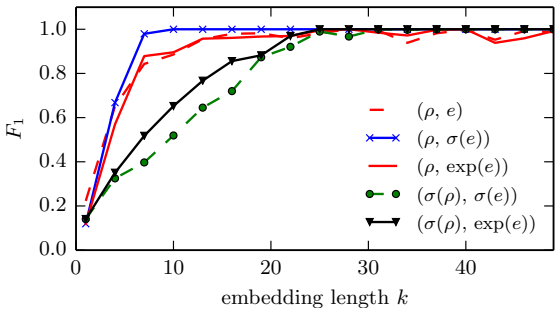


Figure 1. Ability of various methods $\big(\boldsymbol{v}(r), \boldsymbol{v}(e)\big)$ to reconstruct binary matrices, on a sampled KB with 50 entities (249 observed entity pairs) and 20 relations.

We propose a method that avoids this problem by incorporating first-order implications directly into relation representations, combining work of Order Embeddings (Vendrov et al., 2016) and Approximately Boolean Representations (Kruszewski et al., 2015). The idea is to map relation and entity-pair representations into well-chosen subspaces in which the implication rules can be expressed as direct regularizers of relation representations without imposing them on entity representations, too.

This abstract summarizes (Demeester et al., 2016b), extended with some insights from our more recent work (Demeester et al., 2016a).

2. Model

The universal schemas approach for relation extraction (Riedel et al., 2013) applies matrix factorization techniques to make predictions on the validity of facts, given a set $\mathcal{P}$ of entity tuples, a set $\mathcal{R}$ of relations, and

a training set of observed facts. A k-dimensional vector representation is constructed for each entity tuple and relation, and the probability that a particular relation $r \in \mathcal{R}$ and entity tuple $e \in \mathcal{P}$ form a valid fact is modeled as the sigmoid of the dot product of their representations $\boldsymbol{v}(r)$ and $\boldsymbol{v}(e)$:

$$p\big(z = 1|\boldsymbol{v}(r), \boldsymbol{v}(e)\big) = \sigma\big(\boldsymbol{v}(r)^T \boldsymbol{v}(e)\big), \qquad (1)$$

with the binary target variable z indicating validity of the considered fact. The representations $\boldsymbol{v}(r)$ and $\boldsymbol{v}(e)$ can be found by minimizing the logistic loss $\mathcal{L}_F$ over the training facts, together with a set of sampled negative facts.

If we restrict the embeddings in an appropriate way, they can gain useful properties. In particular, we want to inject implications $r_p \Rightarrow r_q$, such that relation r_p always implies relation r_q. Thus we require that, $\forall e \in \mathcal{P}$:

$$p\big(z = 1|\boldsymbol{v}(r_p), \boldsymbol{v}(e)\big) \leq p\big(z = 1|\boldsymbol{v}(r_q), \boldsymbol{v}(e)\big). \qquad (2)$$

If the component-wise inequality $\boldsymbol{v}(r_p) \leq \boldsymbol{v}(r_q)$ holds, and we restrict all components of $\boldsymbol{v}(e)$ to be non-negative, then by construction $\boldsymbol{v}(r_p)^T \boldsymbol{v}(e) \leq \boldsymbol{v}(r_q)^T \boldsymbol{v}(e)$, and with eq. (1), the above requirement is satisfied, whereby the implication rule is no longer grounded in the training facts. This condition can be imposed by introducing an additional loss term $\mathcal{L}_I$ for each implication rule, which penalizes violations of the inequality $\boldsymbol{v}(r_p) \leq \boldsymbol{v}(r_q)$. For details, see (Demeester et al., 2016a). In order to impose the required restriction $\boldsymbol{v}(e) \in \mathbb{R}^{k,+}$ in entity tuple space, we can construct the embeddings by mapping a real-valued vector $\boldsymbol{e} \in \mathbb{R}$ into non-negative space, e.g., as $\boldsymbol{v}(e) = \exp(\boldsymbol{e})$ or $\boldsymbol{v}(e) = \sigma(\boldsymbol{e})$. In line with the work on approximately boolean representations (Kruszewski et al., 2015), we can also try to map the relation embeddings into the k-dimensional hypercube, by transforming the real-valued vector $\boldsymbol{\rho} \in \mathbb{R}$ into the approximately Boolean form $\sigma(\boldsymbol{\rho})$, using the sigmoid function.

3. Experimental Results

We consider an artificial dataset with 20 relations and 50 entities, and calculate the F_1 score for reconstructing all training facts, assuming that all unobserved facts are negative. Fig. 1 shows the result for different combinations of restricting the relation and entity pair embedding spaces, as a function of the embedding size k. The strong restriction of approximately Boolean relation embeddings requires a higher embedding size than with non-restricted relation embeddings. Surprisingly, the model's expressiveness for the case $(\boldsymbol{\rho}, \sigma(\boldsymbol{e}))$ seems not to be reduced as opposed to the unrestricted case $(\boldsymbol{\rho}, \boldsymbol{e})$, although it satisfies the restriction $\boldsymbol{v}(e) \in \mathbb{R}^{k,+}$ required to inject implications. We can therefore step away from the approximately Boolean relation embeddings setting $\sigma(\boldsymbol{\rho}, \exp(\boldsymbol{e}))$ used in (Demeester et al., 2016b). For the following experiment, the transformation $(\boldsymbol{\rho}, \sigma(\boldsymbol{e}))$ is used.

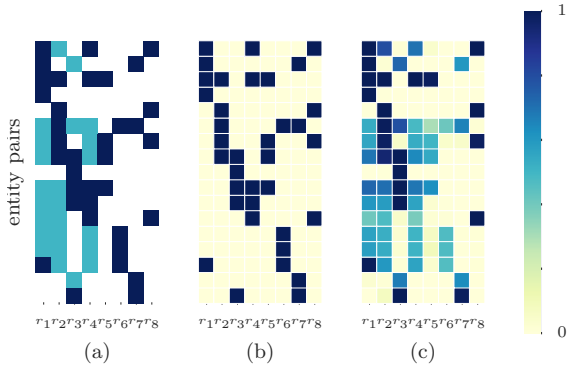


Figure 2. Toy example with 8 relations, 17 observed entity pairs, and 5 implication rules ($r_4 \Rightarrow r_1$), ($r_7 \Rightarrow r_3$), ($r_4 \Rightarrow r_2$), ($r_6 \Rightarrow r_4$), and ($r_5 \Rightarrow r_4$). (a) Original knowledge base (dark: known facts; white: unknown facts; shaded: inferred facts from rules); (b) reconstructed with embedding size 15 with $\mathcal{L} = \mathcal{L}_F$, and (c) with $\mathcal{L} = \mathcal{L}_F + \mathcal{L}_I$.

To visualize what happens when regularizing relation representations based on given implications, we sample a small KB with 8 relations, 17 entity pair observations and the following five implications: $r_4 \Rightarrow r_1$, $r_7 \Rightarrow r_3$, $r_4 \Rightarrow r_2$, $r_6 \Rightarrow r_4$ and $r_5 \Rightarrow r_4$. We add 20% of the facts that can be inferred from these rules as training data and the rest as test facts. Fig. 2(a) shows observed facts (dark blue), as well as test facts (light blue). With $k = 15$, $\mathcal{L}_F$ is able to reconstruct the training data, as shown in Fig. 2(b), but therefore overfits. In contrast, when imposing implications we can reconstruct the training facts and simultaneously infer the test facts implied by these rules (Fig. 2(c)).

4. Conclusion and Future Work

We have presented a new method to incorporate first-order implications into relation representations for knowledge base inference. Our method is scalable, as it alleviates the need for propositionalization of such formulae.

This abstract summarizes the results from (Demeester et al., 2016b). More recent experimental results on the New York Times dataset used in (Rocktäschel et al., 2015), confirm that the presented method scales to larger problems, and can be used to inject many formulae extracted from ontologies. This is described in detail in (Demeester et al., 2016a).

References

Demeester, T., Rocktäschel, T., & Riedel, S. (2016a). Lifted rule injection for relation embeddings. *arXiv preprint arXiv:1606.08359, accepted at EMNLP 2016*.

Demeester, T., Rocktäschel, T., & Riedel, S. (2016b). Regularizing relation representations by first-order implications. *Proceedings of the 5th Workshop on Automated Knowledge Base Construction (AKBC)*.

Kruszewski, G., Paperno, D., & Baroni, M. (2015). Deriving boolean structures from distributional vectors. *Transactions of the Association for Computational Linguistics*, *3*, 375–388.

Miller, G. A. (1995). Wordnet: a lexical database for english. *Communications of the ACM*, *38*, 39–41.

Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2015). A review of relational machine learning for knowledge graphs: From multi-relational link prediction to automated knowledge graph construction. *arXiv preprint arXiv:1503.00759*.

Riedel, S., Yao, L., McCallum, A., & Marlin, B. M. (2013). Relation extraction with matrix factorization and universal schemas.

Rocktäschel, T., Singh, S., & Riedel, S. (2015). Injecting Logical Background Knowledge into Embeddings for Relation Extraction. *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*.

Vendrov, I., Kiros, R., Fidler, S., & Urtasun, R. (2016). Order-embeddings of images and language. *arXiv preprint*, *abs/1511.06361*.

Wang, Q., Wang, B., & Guo, L. (2015). Knowledge base completion using embeddings and rules. *Proceedings of the 24th International Conference on Artificial Intelligence* (pp. 1859–1865). Buenos Aires, Argentina: AAAI Press.