Learning Multi-faceted Knowledge Graph Embeddings for Natural Language Processing

Muhao Chen, Carlo Zaniolo
Department of Computer Science, UCLA
{muhaochen, zaniolo}@cs.ucla.edu

Abstract

Knowledge graphs have challenged the existing embedding-based approaches for representing their multifacetedness. To address some of the issues, we have investigated some novel approaches that (i) capture the multilingual transitions on different language-specific versions of knowledge, and (ii) encode the commonly existing monolingual knowledge with important relational properties and hierarchies. In addition, we propose the use of our approaches in a wide spectrum of NLP tasks that have not been well explored by related works.

1 Introduction

Knowledge graph (KG) embeddings are the essential tools that transfer the important concepts of the complicated human languages into machine-understandable representations. The past half decade of research has paid much attention to translation-based methods, which provide simple techniques to encode entities in low-dimensional spaces and capture relations as means of translations among entity vectors.

Although efforts have been made to improve such methods by introducing various forms of relation-specific entity projections (see [Chen *et al.*, 2017] for summary), almost all works focus on characterizing monolingual knowledge defined on simple or multi-mapping relations. This is far from enough for completely representing the KGs, inasmuch as KGs usually contain more complicated forms of knowledge, including multilingual knowledge that synchronizes multiple versions of KGs [Mahdisoltani *et al.*, 2015], as well as monolingual knowledge that enforces relational properties such as transitivity and symmetry, and forms hierarchies [McGuinness *et al.*, 2004]. We propose some embedding approaches that pursue to joint learning of monolingual and multilingual knowledge, and better preserving relational properties and hierarchies of monolingual knowledge in learning process. So far as related works largely limit the use of KG embeddings to tasks directly associated with KGs [Ji *et al.*, 2016], like relation extraction and triple classification, we propose the use of our approaches in a wider spectrum of NLP tasks.

2 Proposed Approaches

We give the formalization of the corpora, with the highlight on the multi-faceted knowledge we consider.

2.1 Formalization of Corpora

In a knowledge base KB, we use $\mathcal{L}$ to denote the set of languages, and $\mathcal{L}^2$ to denote the *unordered* language pairs. For $L \in \mathcal{L}$, G_L denotes the language-specific KG of L, and E_L and R_L respectively denote the corresponding vocabularies of entities and relations. $T = (h, r, t)$ denotes a triple in G_L such that $h, t \in E_L$ and $r \in R_L$. Boldfaced $\mathbf{h}$, $\mathbf{r}$, $\mathbf{t}$ represent the embedding vectors of head h, relation r, and tail t. For a language pair $(L_1, L_2) \in \mathcal{L}^2$, $\delta(L_1, L_2)$ denotes the alignment set which contains the aligned pairs of triples.

Within a G_L, we extend the relations by $R_L = R_{tr} \cup R_s \cup R_h \cup R_o$, which respectively denote the sets of transitive, symmetric, hierarchical, and other simple relations. Thereof, $R_h = R_r \cup R_c$ where R_r denotes refinement relations that divide entities to finer ones, and R_c denotes coercion relations that merge entities to coarser ones [Chen *et al.*, 2016].

2.2 Learning Multilingual Knowledge

Leveraging KG embeddings to multilingual knowledge no doubt provides more generic representations that benefits cross-lingual NLP. But it is non-trivial for several reasons: (i) multilingual knowledge has far larger domains than monolingual relations; (ii) it applies to both entities and relations with incoherent vocabularies in every language; (iii) usually only a small portion of KB has been aligned with such knowledge.

We have proposed the *MTransE* model [Chen *et al.*, 2017] that employs two model components to learn on the two facets of KB: *knowledge model* that encodes the entities and relations from each language-specific graph structure, and *alignment model* that learns cross-lingual transitions. It is noteworthy that, MTransE is defined on a pair of language. A set of model for each $(L_i, L_j) \in \mathcal{L}^2$ composes the solution for more than two languages w.l.o.g.

Knowledge Model. For each language $L \in \mathcal{L}$, a dedicated embedding space $\mathbb{R}^k_L$ is assigned for vectors of E_L and R_L, where $\mathbb{R}$ is the field of real numbers. We employ TransE [Bordes *et al.*, 2013] for each language by adopting the objective function $S_K = \sum_{L \in \{L_i, L_j\}} \sum_{(h,r,t) \in G_L} \|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$, which benefits the cross-lingual NLP tasks with the uniform representation of entities under different contexts of relations.

Alignment Model. Alignment model constructs the transitions between the vector spaces of L_i and L_j according to the objective function $S_A = \sum_{(T,T') \in \delta(L_i, L_j)} S_a(T, T')$. The alignment score $S_a(T, T')$ thereof iterates through all pairs of aligned triples. Three different techniques to represent the alignment scores are considered, i.e., distance-based axis calibration, translation vectors, and linear transforms. These techniques lead to five forms of S_a.

MTransE minimizes the objective function $J = S_K + \alpha S_A$ via jointly training on triples and alignment sets, where α is a hyperparameter that weights between two model components. Several variants of MTransE are implemented based on different techniques applied in the alignment model.

2.3 Learning with Properties and Hierarchies

In many KGs, monolingual knowledge commonly enforces relational properties, and forms hierarchies. For example, 85% of triples in ConceptNet [Speer and others, 2013] and 96% in Yago3 [Mahdisoltani *et al.*, 2015] are with relational properties, while 60% and 38% of their triples are with hierarchical relations respectively. Thus, we investigate another model *On2Vec* that forthputs the *Component-specific Model* and *Hierarchy Model* to handle such knowledge.

Component-specific Model. Existing approaches mainly apply the same relation-specific projection on both head and tail entities [Ji *et al.*, 2016]. This inevitably causes conflicts on triples with relational properties:

- Consider $r \in R_{tr}$ and (e_1, r, e_2), (e_2, r, e_3), $(e_1, r, e_3) \in G_L$, where e_1, e_2, and e_3 are projected to $\mathbf{e_{1r}}$, $\mathbf{e_{2r}}$, and $\mathbf{e_{3r}}$ in the relation space of r. Since $\mathbf{e_{1r}} + \mathbf{r} \approx \mathbf{e_{2r}}$ and $\mathbf{e_{2r}} + \mathbf{r} \approx \mathbf{e_{3r}}$ hold for the first and second triples, it is impossible for $\mathbf{e_{1r}} + \mathbf{r} \approx \mathbf{e_{3r}}$ to hold as $\mathbf{r}$ is nonzero.
- Consider $r \in R_s$ and (e_1, r, e_2), $(e_2, r, e_1) \in G_L$. It is impossible for both $\mathbf{e_{1r}} + \mathbf{r} \approx \mathbf{e_{2r}}$ and $\mathbf{e_{2r}} + \mathbf{r} \approx \mathbf{e_{1r}}$ to hold, since $\mathbf{r}$ is nonzero.

To eliminate such conflicts, the component-specific model applies two different relation-specific projections $f_{1,r}$ and $f_{2,r}$ on head and tail entities respectively to the same relation space. $f_{1,r}$ and $f_{2,r}$ can be implemented as hyperplane mappings [Wang *et al.*, 2014] or linear transforms.

Hierarchy Model. Given $r \in R_h$, we define $\sigma(e, r)$ as the refine operator to fetch the finer entities that apply r with the coarser entity e. Hierarchy model serves as an auxiliary learning process for R_h towards the goal of converging the projected embeddings of $e' \in \sigma(e, r)$ within a tight neighborhood given each e and r, while penalizing other unrelated entities that fall inside the margin.

Training of On2Vec jointly optimizes the objective of both model components. Bernoulli negative sampling [Wang *et al.*, 2014] is adopted for more efficient learning. Experimental results on relation extraction show that On2Vec outperforms the related approaches [Chen, 2017], which verifies the effectiveness of the two On2Vec model components.

3 NLP Applications

We propose to apply these approaches to several NLP tasks.

Knowledge Alignment. We have applied MTransE to both entity and triple-wise cross-lingual alignment tasks on large KGs derived from Wikipedia and ConceptNet, which received promising results and outperformed approaches based on multilingual word embeddings [Chen *et al.*, 2017]. Results of monolingual tasks also show that MTransE better preserves monolingual knowledge than its monolingual counterpart.

Semantics Relatedness Analysis (SRA). Unsupervised SRA tasks such as LP50 [Chen, 2017] request machine-estimated document semantic distances to receive a high Pearson's correlation with human judgement. A straightforward solution is via vector representations of documents by aggregating the embeddings of Wikipedia entities that are recognized in each document. Currently, as we use Annotated Skip-gram trained on Wikipedia articles to conceive the vectors, this simple solution is effective enough to outperform many approaches [Chen, 2017]. As next step, we will represent the documents by training On2Vec on the Wikipedia-derived Yago3. Similarly, we will use MTransE to enable cross-lingual SRA.

Open Information Extraction (OpenIE). Neural openIE [Lin *et al.*, 2016] relies on word embeddings to obtain the tensor representations of sentences. A practical alternative is to use word-knowledge joint embeddings [Chen, 2017] based on On2Vec and Annotated Skip-gram, which stands on the assumption that KG embeddings better highlight the important concepts in sentences [Ji *et al.*, 2016], amongst which the neural classier aims at predicting the relations.

Sentiment Analysis. Deep neural networks similar to the one for openIE will be used to tackle sentiment analysis tasks summarized in [Kim, 2014]. We will evaluate the following sentence representation techniques: (i) joint embedding-based tensor representations; (ii) aggregated embeddings of highlighted KG entities; (iii) combination of (i) and (ii).

References

[Bordes *et al.*, 2013] Antoine Bordes, Nicolas Usunier, et al. Translating embeddings for modeling multi-relational data. In *NIPS*, 2013.

[Chen *et al.*, 2016] Muhao Chen, Shi Gao, et al. Converting spatiotemporal data among heterogeneous granularity systems. In *FUZZ-IEEE*, 2016.

[Chen *et al.*, 2017] Muhao Chen, Yingtao Tian, et al. Multilingual knowledge graph embeddings for cross-lingual knowledge alignment. *IJCAI*, 2017.

[Chen, 2017] Muhao Chen. Learning knowledge graph embeddings for natural language process. In *UCLA CSD Tech Report*, number 170002, 2017.

[Ji *et al.*, 2016] Guoliang Ji, Kang Liu, et al. Knowledge graph completion with adaptive sparse transfer matrix. In *AAAI*, 2016.

[Kim, 2014] Yoon Kim. Convolutional neural networks for sentence classification. *EMNLP*, 2014.

[Lin *et al.*, 2016] Yankai Lin, Shiqi Shen, et al. Neural relation extraction with selective attention over instances. In *ACL*, 2016.

[Mahdisoltani *et al.*, 2015] Farzaneh Mahdisoltani, Joanna Biega, et al. Yago3: A knowledge base from multilingual Wikipedias. In *CIDR*, 2015.

[McGuinness *et al.*, 2004] Deborah L McGuinness, Frank Van Harmelen, et al. Owl web ontology language overview. *W3C recommendation*, 10(10), 2004.

[Speer and others, 2013] Robert Speer et al. Conceptnet 5: A large semantic network for relational knowledge. *The People's Web Meets NLP*, 2013.

[Wang *et al.*, 2014] Zhen Wang, Jianwen Zhang, et al. Knowledge graph embedding by translating on hyperplanes. In *AAAI*, 2014.