A Convolutional Neural Network-based Model for Knowledge Base Completion

Dat Quoc Nguyen

Joint work with:
Dai Quoc Nguyen, Tu Dinh Nguyen and Dinh Phung

April 16, 2018
Introduction

• Word vectors learned from a large corpus can model relational similarities or linguistic regularities between pairs of words as translations in the projected vector space:

$$\mathbf{v}_{king} - \mathbf{v}_{man} \approx \mathbf{v}_{queen} - \mathbf{v}_{woman}$$

A “royal” relationship between “king” and “man”, and between “queen” and “woman”
Introduction

- Let consider the country and capital pairs to be pairs of entities rather than word types
  - We represent country and capital entities by low-dimensional and dense vectors
  - The relational similarity between word pairs is presumably to capture a "is capital of" relationship between country and capital entities
  - We also represent this relationship by a translation vector in the entity vector space

\[ \mathbf{v}_{Tokyo} + \mathbf{v}_{is\_capital\_of} - \mathbf{v}_{Japan} \approx 0 \]
\[ \mathbf{v}_{Berlin} + \mathbf{v}_{is\_capital\_of} - \mathbf{v}_{Germany} \approx 0 \]
\[ \mathbf{v}_{Rome} + \mathbf{v}_{is\_capital\_of} - \mathbf{v}_{Italy} \approx 0 \]
\[ \mathbf{v}_{Lisbon} + \mathbf{v}_{is\_capital\_of} - \mathbf{v}_{Portugal} \approx 0 \]
Introduction

- This intuition inspired the TransE model—a well-known embedding model for **KB completion** or link prediction in KBs (Bordes et al., 2013)

\[
\begin{align*}
\mathbf{v}_{Tokyo} + \mathbf{v}_{is\text{-}capital\_of} - \mathbf{v}_{Japan} & \approx 0 \\
\mathbf{v}_{Berlin} + \mathbf{v}_{is\text{-}capital\_of} - \mathbf{v}_{Germany} & \approx 0 \\
\mathbf{v}_{Rome} + \mathbf{v}_{is\text{-}capital\_of} - \mathbf{v}_{Italy} & \approx 0 \\
\mathbf{v}_{Lisbon} + \mathbf{v}_{is\text{-}capital\_of} - \mathbf{v}_{Portugal} & \approx 0
\end{align*}
\]
• KBs are collections of real-world triples, where each triple or fact \((h, r, t)\) represents some relation \(r\) between a head entity \(h\) and a tail entity \(t\)
  • Entities are real-world things or objects such as persons, places, organizations, music tracks or movies
  • Each relation type defines a certain relationship between entities
  • E.g., the relation type “\(\text{child of}\)” relates person entities with each other, while the relation type “\(\text{born in}\)” relates person entities with place entities
• KBs thus are useful resources for many NLP tasks
Introduction

• Issue: KBs are far from complete
  • In English DBpedia 2014, 60% of person entities miss a place of birth and 58% of the scientists do not have a fact about what they are known for (Krompaß et al., 2015)
  • In Freebase, 71% of 3 million person entities miss a place of birth, 75% do not have a nationality while 94% have no facts about their parents (West et al., 2014)

• So, a question answering application based on an incomplete KB would not provide a correct answer given a correctly interpreted question
  • It would be impossible to answer the question “where was Jane born?”
Introduction

• KB completion or Link prediction:
  • Predict whether a relationship/triple not in the KB is likely to be true, i.e., to add new triples by leveraging existing triples in the KB
  • E.g., Predict the missing tail entity in the incomplete triple (Jane, born in, ?) or predict whether the triple (Jane, born in, Miami) is correct or not
Embedding models for KB completion

- Embedding models for KB completion have been proven to give state-of-the-art link prediction performances
  - Entities are represented by latent feature vectors
  - Relation types are represented by latent feature vectors, matrices or third-order tensors
Embedding models for KB completion

• For each triple \((h, r, t)\), the embedding models define a score function \(f(h; r; t)\) of its implausibility:
  
  • Choose \(f\) such that the score \(f(h, r, t)\) of a correct triple \((h, r, t)\) is smaller than the score \(f(h', r', t')\) of an incorrect triple \((h', r', t')\)

• TransE:

\[
\begin{align*}
    f_{\text{TransE}}(h, r, t) &= \| v_h + v_r - v_t \|_{\ell_1/2} \\
    \| v_{\text{Tokyo}} + v_{\text{is\_capital\_of}} - v_{\text{Japan}} \|_{\ell_1/2} &< \| v_{\text{Tokyo}} + v_{\text{is\_capital\_of}} - v_{\text{Portugal}} \|_{\ell_1/2}
\end{align*}
\]
### Embedding models for KB completion

<table>
<thead>
<tr>
<th>Model</th>
<th>Score function $f(h, r, t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unstructured</td>
<td>$|v_h - v_t|<em>{\ell</em>{1/2}}$</td>
</tr>
<tr>
<td>SE</td>
<td>$|W_{r,1}v_h - W_{r,2}v_t|<em>{\ell</em>{1/2}}; W_{r,1}, W_{r,2} \in \mathbb{R}^{k \times k}$</td>
</tr>
<tr>
<td>TransE</td>
<td>$|v_h + v_r - v_t|<em>{\ell</em>{1/2}}; v_r \in \mathbb{R}^k$</td>
</tr>
<tr>
<td>STransE</td>
<td>$|W_{r,1}v_h + v_r - W_{r,2}v_t|<em>{\ell</em>{1/2}}; W_{r,1}, W_{r,2} \in \mathbb{R}^{k \times k}; v_r \in \mathbb{R}^k$</td>
</tr>
<tr>
<td>DISTMULT</td>
<td>$v_h^T W_r v_t; W_r$ is a diagonal matrix $\in \mathbb{R}^{k \times k}$</td>
</tr>
<tr>
<td>Bilinear-COMP</td>
<td>$v_h^T W_{r_1} W_{r_2} ... W_{r_m} v_t; W_{r_1}, W_{r_2}, ..., W_{r_m} \in \mathbb{R}^{k \times k}$</td>
</tr>
<tr>
<td>TransE-COMP</td>
<td>$|v_h + v_{r_1} + v_{r_2} + ... + v_{r_m} - v_t|<em>{\ell</em>{1/2}}; v_{r_1}, v_{r_2}, ..., v_{r_m} \in \mathbb{R}^k$</td>
</tr>
<tr>
<td>ConvE</td>
<td>$v_t^T g(\text{vec}(g(\text{concat}(\overline{v}_h, \overline{v}_r) * \Omega)) W); g$ denotes a non-linear function</td>
</tr>
<tr>
<td>Our ConvKB</td>
<td>$w^T \text{concat}(g([v_h, v_r, v_t] * \Omega)); *$ denotes a convolution operator</td>
</tr>
</tbody>
</table>
Embedding models for KB completion

• A common objective function is the following margin-based function:

\[ \mathcal{L} = \sum_{(h, r, t) \in \mathcal{G}} \max(0, \gamma + f(h, r, t) - f(h', r, t')) \]

• \( \gamma \) is the margin hyper-parameter

• \( \mathcal{G}'_{(h, r, t)} \) is the set of incorrect triples generated by corrupting the correct triple \( (h, r, t) \in \mathcal{G} \)
A CNN-based model for KB completion

- Each embedding triple \((v_h, v_r, v_t)\) are viewed as a 3-column matrix:
  \[
  A = [v_h, v_r, v_t]
  \]
- Each filter \(\omega \in \mathbb{R}^{1 \times 3}\) is repeatedly operated over every row of \(A\) to generate a feature map:
  \[
  v = [v_1, v_2, \ldots, v_k] \text{ with } v_i = g(\omega \cdot A_{i,:} + b)
  \]
- Let \(\Omega\) and \(*\) denote the set of filters and a convolution operator, respectively.
- Our ConvKB formally defines a score function as:
  \[
  f(h, r, t) = \text{concat}(g([v_h, v_r, v_t] \ast \Omega)) \cdot w
  \]
A CNN-based model for KB completion

- ConvKB formally defines a score function as
  \[ f(h, r, t) = \text{concat} \left( g \left( [v_h, v_r, v_t] * \Omega \right) \right) \cdot w \]
- If we only use one filter with \( g \) be the vector norm and fix \( \omega = [1, 1, -1] \) and \( w = 1 \) during training, then ConvKB reduces to the plain TransE
  - ConvKB model can be viewed as an extension of TransE
KB completion experiments

On both FB13 and WN11, each validation and test set also contains the same number of incorrect triples as the number of correct triples.

| Dataset  | |E| | |R| | #Triples in train/valid/test |
|----------|------|------|----------------------|
| FB15k-237| 14,541 | 237 | 272,115 17,535 20,466 |
| WN18RR   | 40,943 | 11  | 86,835  3,034 3,134 |
| FB13     | 75,043 | 13  | 316,232 5,908 23,733 |
| WN11     | 38,696 | 11  | 112,581 2,609 10,544 |

- **Link prediction task:**
  - Predict h given (? , r , t) or predict t given (h , r , ?) where ? denotes the missing element
  - Corrupt each correct test triple (h, r, t) by replacing either h or t by each of the possible entities
  - Rank these candidates by their implausibility value
  - Metrics: mean rank (MR), mean reciprocal rank (MRR), and Hits@10 (i.e., the proportion of the valid test triples ranking in top 10 predictions)
KB completion experiments

- Link prediction results:

<table>
<thead>
<tr>
<th>Method</th>
<th>WN18RR</th>
<th>FB15k-237</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MR</td>
<td>MRR</td>
</tr>
<tr>
<td>IRN (Shen et al., 2017)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>KBGAN (Cai and Wang, 2018)</td>
<td>–</td>
<td>0.213</td>
</tr>
<tr>
<td>DISTMULT (Yang et al., 2015) [*]</td>
<td>5110</td>
<td>0.43</td>
</tr>
<tr>
<td>ComplEx (Trouillon et al., 2016) [*]</td>
<td>5261</td>
<td>0.44</td>
</tr>
<tr>
<td>ConvE (Dettmers et al., 2018)</td>
<td>5277</td>
<td><strong>0.46</strong></td>
</tr>
<tr>
<td>TransE (Bordes et al., 2013) (our results)</td>
<td>3384</td>
<td>0.226</td>
</tr>
<tr>
<td>Our ConvKB model</td>
<td><strong>2554</strong></td>
<td>0.248</td>
</tr>
<tr>
<td>$KB_{LRN}$ (García-Durán and Niepert, 2017)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>R-GCN+ (Schlichtkrull et al., 2017)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Neural LP (Yang et al., 2017)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Node+LinkFeat (Toutanova and Chen, 2015)</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
KB completion experiments

- Hits@10 on FB15k-237:

Predicting head:
- 1-1: 53.6, 52.6
- 1-M: 58.9, 58.3
- M-1: 9.9, 38.6
- M-M: 39.8, 47.5

Predicting tail:
- 1-1: 53.1, 49.5
- 1-M: 83.9, 84.3
- M-1: 7.2, 9.1
- M-M: 55.1, 53.6
KB completion experiments

- Hits@10 on WN18RR:
KB completion experiments

- **Triple classification task:**
  - Predict whether a triple \((h, r, t)\) is correct or not
  - Set a relation-specific threshold \(\theta_r\) for each relation type \(r\)
  - For an unseen test triple \((h, r, t)\), if \(f(h, r, t)\) is smaller than \(\theta_r\) then the triple will be classified as correct, otherwise incorrect
  - Relation-specific thresholds are determined by maximizing the micro-averaged accuracy on the validation set

<table>
<thead>
<tr>
<th>Method</th>
<th>WN11</th>
<th>FB13</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTN [41]</td>
<td>70.6</td>
<td>87.2</td>
<td>78.9</td>
</tr>
<tr>
<td>TransH [53]</td>
<td>78.8</td>
<td>83.3</td>
<td>81.1</td>
</tr>
<tr>
<td>TransR [27]</td>
<td>85.9</td>
<td>82.5</td>
<td>84.2</td>
</tr>
<tr>
<td>TransD [19]</td>
<td>86.4</td>
<td>89.1</td>
<td>87.8</td>
</tr>
<tr>
<td>TransR-FT [12]</td>
<td>86.6</td>
<td>82.9</td>
<td>84.8</td>
</tr>
<tr>
<td>TranSparse-S [20]</td>
<td>86.4</td>
<td>88.2</td>
<td>87.3</td>
</tr>
<tr>
<td>TranSparse-US [20]</td>
<td>86.8</td>
<td>87.5</td>
<td>87.2</td>
</tr>
<tr>
<td>ManifoldE [56]</td>
<td>87.5</td>
<td>87.2</td>
<td>87.4</td>
</tr>
<tr>
<td>TransG [57]</td>
<td>87.4</td>
<td>87.3</td>
<td>87.4</td>
</tr>
<tr>
<td>lppTransD [64]</td>
<td>86.2</td>
<td>88.6</td>
<td>87.4</td>
</tr>
<tr>
<td>TransE [5] (our results)</td>
<td>86.5</td>
<td>87.5</td>
<td>87.0</td>
</tr>
<tr>
<td>Our ConvKB model</td>
<td><strong>87.6</strong></td>
<td><strong>88.8</strong></td>
<td><strong>88.2</strong></td>
</tr>
<tr>
<td>TransE-NMM [33]</td>
<td>86.8</td>
<td>88.6</td>
<td>87.7</td>
</tr>
<tr>
<td>TEKE_H [52]</td>
<td>84.8</td>
<td>84.2</td>
<td>84.5</td>
</tr>
<tr>
<td>Bilinear-COMP [16]</td>
<td>77.6</td>
<td>86.1</td>
<td>81.9</td>
</tr>
<tr>
<td>TransE-COMP [16]</td>
<td>80.3</td>
<td>87.6</td>
<td>84.0</td>
</tr>
</tbody>
</table>
KB completion experiments

• Triple classification results on FB13
Conclusion

• We have proposed an embedding model ConvKB for knowledge base completion

• ConvKB outperforms previous state-of-the-art models:
  • On two benchmark link prediction datasets WN18RR and FB15k-237
  • On two benchmark triple classification datasets WN11 and FB13

• Code: https://github.com/daiquocnguyen/ConvKB

• References: