Iterative Entity Alignment via Joint Knowledge Embeddings

Hao Zhu
Jointly work with Ruobing Xie, Zhiyuan Liu, Maosong Sun

Natural Language Processing Lab,
Department of Computer Science and Technology,
Tsinghua University
Entity Alignment

• Goal
  – Align synonymous entity pairs from heterogeneous Knowledge Graphs
BACKGROUND

Knowledge Graphs
Knowledge Representation Learning
Knowledge Alignment
Existing Knowledge Graphs
Knowledge Alignment

- Graph-based models
  - time-consuming on large-scale KGs
- Other conventional models
  - crowd-sourcing
  - well-designed hand-crafted features
- MTransE [Chen et al. 2016]
  - similar idea, but from different assumption
  - experimental result shows our method is better
Knowledge Representation Learning

- TransE [Bordes et al., 2013] and its extensions
- RESCAL [Nickel et al., 2011; 2012]
- HOLE [Nickel et al., 2016]
- NTN [Socher et al., 2013]
TransE

- Embedding:
  - Entity: vectors
  - Relation: translation vectors

- Goal: $h + r = t$
PTransE [Lin et. al 2015]

• Besides entities and relations, also embed relation path into the same space.
What could KRL help us?

Representation \quad \leftrightarrow \quad \text{Intrinsic Meaning}

Closer Representations \quad \leftrightarrow \quad \text{Higher Probability to be Synonymous}
Knowledge Representation Learning
Parameter Sharing Model
Iterative Alignment Model

Our Model (ITransE)
Knowledge Representation Learning

Knowledge Graph

Alignment Seeds

Knowledge Graph 1 (KG1)
- e1
- e2
- e3
- r2

Knowledge Graph 2 (KG2)
- e1
- e2

Relationship Among Embeddings

\[ e_2 + r_2 \sim e_3 \]

Reliability = 1

NLP Lab
Tsinghua University
Parameter Sharing Model

Knowledge Graph

Alignment Seeds

KG₁

KG₂

Relationship Among Embeddings

Reliability = 1

NLP Lab
Tsinghua University
Iterative learning Model

Alignment Seeds

Newly Aligned Entity Pairs

Knowledge Graph

Relationship Among Embeddings

Reliability = 1

Reliability = R(●, ●)
Empirical Evaluation

• Data
  – DFB-1,2,3

<table>
<thead>
<tr>
<th>Dataset</th>
<th>R</th>
<th>E</th>
<th>$T_1$</th>
<th>$T_2$</th>
<th>L</th>
<th>#Valid</th>
<th>O</th>
</tr>
</thead>
<tbody>
<tr>
<td>DFB-1</td>
<td>1,345</td>
<td>14,951</td>
<td>444,159</td>
<td>444,160</td>
<td>5,000</td>
<td>1,000</td>
<td>0.5</td>
</tr>
<tr>
<td>DFB-2</td>
<td>1,345</td>
<td>14,951</td>
<td>444,159</td>
<td>444,160</td>
<td>500</td>
<td>1,000</td>
<td>0.5</td>
</tr>
<tr>
<td>DFB-3</td>
<td>1,345</td>
<td>14,951</td>
<td>325,717</td>
<td>325,717</td>
<td>500</td>
<td>1,000</td>
<td>0.1</td>
</tr>
</tbody>
</table>

  – DFB-4: Training set, test set and auxiliary training set are 399, 856/59, 071/399, 857 respectively.

• Task
  – Entity Alignment
  – Knowledge Completion
## Entity Alignment

- **Goal**
  - infer the synonymous entity pairs

<table>
<thead>
<tr>
<th>Metric</th>
<th>DFB-1</th>
<th>DFB-2</th>
<th>DFB-3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hits@1</td>
<td>Hits@10</td>
<td>Mean Rank</td>
</tr>
<tr>
<td>MTransE (LT)</td>
<td>38.9</td>
<td>61.0</td>
<td>237.7</td>
</tr>
<tr>
<td>MTransE (TB)</td>
<td>13.6</td>
<td>35.1</td>
<td>547.7</td>
</tr>
<tr>
<td>TransE + PS</td>
<td>61.9</td>
<td>79.2</td>
<td>105.2</td>
</tr>
<tr>
<td>ITransE (HA)</td>
<td>62.6</td>
<td>78.9</td>
<td>100.0</td>
</tr>
<tr>
<td>ITransE (SA)</td>
<td><strong>67.1</strong></td>
<td><strong>83.1</strong></td>
<td><strong>80.1</strong></td>
</tr>
<tr>
<td>PTransE + PS</td>
<td>65.8</td>
<td>83.4</td>
<td>62.9</td>
</tr>
<tr>
<td>IPTransE (HA)</td>
<td>66.1</td>
<td>83.3</td>
<td>59.1</td>
</tr>
<tr>
<td>IPTransE (SA)</td>
<td><strong>71.7</strong></td>
<td><strong>86.5</strong></td>
<td><strong>49.0</strong></td>
</tr>
</tbody>
</table>
Hits@1 and Mean Rank of our methods through different iterations. (Hits@10 has similar trends to Hits@1.) We conduct soft alignment every 500 iterations from the 1000-th iteration.
Knowledge Completion

- **Goal**
  - help learn better knowledge embeddings

<table>
<thead>
<tr>
<th>Metric</th>
<th>Mean Rank</th>
<th>Entity Prediction</th>
<th>Relation Prediction</th>
<th>Hits@1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw</td>
<td>Filter</td>
<td>Raw</td>
<td>Filter</td>
</tr>
<tr>
<td>MTransE (LT)</td>
<td>240.8</td>
<td>131.3</td>
<td>36.4</td>
<td>47.3</td>
</tr>
<tr>
<td>MTransE (TB)</td>
<td>851.3</td>
<td>759.7</td>
<td>9.4</td>
<td>10.8</td>
</tr>
<tr>
<td>TransE</td>
<td>246.1</td>
<td>131.6</td>
<td>42.5</td>
<td>54.3</td>
</tr>
<tr>
<td>TransE + Aux</td>
<td>232.8</td>
<td>121.5</td>
<td>43.3</td>
<td>54.9</td>
</tr>
<tr>
<td>ITransE (SA)</td>
<td>209.2</td>
<td>101.0</td>
<td><strong>44.2</strong></td>
<td><strong>55.1</strong></td>
</tr>
<tr>
<td>PTransE</td>
<td>213.0</td>
<td>97.2</td>
<td>50.9</td>
<td>72.1</td>
</tr>
<tr>
<td>PTransE + Aux</td>
<td>206.3</td>
<td>80.4</td>
<td>52.7</td>
<td>80.7</td>
</tr>
<tr>
<td>IPTransE (SA)</td>
<td><strong>197.5</strong></td>
<td><strong>70.6</strong></td>
<td><strong>53.0</strong></td>
<td><strong>80.8</strong></td>
</tr>
</tbody>
</table>

*Note: The numbers indicate performance metrics such as mean rank and hits at a certain threshold.*
Conclusion

• This paper presents iterative entity alignment via joint knowledge embeddings, by encoding both entities and relations of various KGs into a unified semantic space.

• A simple and effective Parameter Sharing Model

• An Iterative Alignment Model

• We evaluate on entity alignment and knowledge graph completion.

• Experiment results show the effectiveness of our methods as compared with other baselines.
Future Work

• incorporate rich external information of KGs for entity alignment
• explore the effectiveness of other KRL models in our methods for entity alignment.

• Our code and data will be available at https://github.com/thunlp/IEAJKE
Questions?