Visual Pivoting for (Unsupervised) Entity Alignment

Fangyu Liu¹, Muhao Chen²,³, Dan Roth³ & Nigel Collier¹

¹ Language Technology Lab, TAL, University of Cambridge, UK
² Viterbi School of Engineering, University of Southern California, USA
³ Department of Computer and Information Science, University of Pennsylvania, USA
Outline

- The task: Entity Alignment
- Status quo: Current methods and challenges
- Method: Vision as a pivot
- Experimental results
The task: Entity Alignment
The Task: Entity Alignment

Entity alignment is the task of linking entities with the same real-world identity from different knowledge graphs (KGs).
Why is Entity Alignment important?

Query: (The Tale of Genji, Genre, ?e)

Novel

Monogatari (story)
Love story
Royal family story
Realistic novel
Ancient literature
Why is Entity Alignment important?
Database merge, e.g. Product & E-commerce graphs
Status quo: current methods
Status quo: Embedding-based methods
Challenge: Knowledge Graphs are sparse
Solution: exploit auxiliary information in knowledge graphs

- entity nodes
- typed relations
- attributes
- descriptions

---

[Yang+, EMNLP-IJCNLP 2019]
Challenges remain

• Performance is still unsatisfactory
  • On DBP15k, Hits@1 ranges from 50% - 60% depending on the language pairs

• Relies on a large seed dictionary (as supervision)
  • For DBP15k, a 4.5k seed dictionary is used, i.e. 30% nodes of the whole graph are annotated for training
Method: Vision as a pivot
Motivation

Images are widely available in knowledge graphs.
Motivation

- Essential elements to KGs, very informative
- Verified by crowd (accurate and no ambiguity)
- Invariant to languages / schemata of the KGs
Motivation

REGULAR KG (ENGLISH)

FAST AND FURIOUS

DWAYNE JOHNSON

starring

CHRIS HEMSWORTH

before

ALF BALDWIN

after

almaMater

UNIVERSITY OF MIAMI

REGULAR KG EMBEDDING SPACE

REGULAR KG (CHINESE)

玩命關頭

巨石強森

演員

語言

英語
Motivation
Method: Multi-modal KG embeddings

Jointly embed (1) graph structures + (2) relations + (3) attributes + (4) images.

- Born: ...
- Citizenship: ...
- Years active: ...
- Debut ...

Diagram:
- Topological features (affinity matrix)
- Relation features (one-hot encodings)
- Attribute features (one-hot encodings)
- Visual features (RGB matrix)

Network:
- GCN
- FC layer
- CNN (ResNet152)
- Concatenation

Multi-modal KG representations
Method: Alignment learning

Neighbouring Component Analysis loss [Liu+, AAAI 2020]:

\[ \mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{\alpha} \log \left( 1 + \sum_{m \neq i} e^{\alpha S_{mi}} \right) + \frac{1}{\alpha} \log \left( 1 + \sum_{n \neq i} e^{\alpha S_{in}} \right) - \log \left( 1 + \beta S_{ii} \right) \right) \]

push negative pairs away  
pull positive pairs together

Modality-specific alignment + joint alignment:

\[ \mathcal{L}_{\text{Joint}} = \sum_{i} \mathcal{L}_{i} + \mathcal{L}_{\text{Multi-modal}} \]

apply to features of each modality  
apply to the concatenated feature
Method: Unsupervised setting

Induce visual seed alignment:

Images from Japanese DBpedia

Images from English DBpedia

0.9 0.2 0.1 0.2

0.2 0.7 0.1 0.5

0.2 0.0 0.8 0.0

0.1 0.1 0.2 0.2

0.9 0.8 0.7 0.5

cut-off
Method: Iterative learning

- **visual seed dictionary**
- **unaligned entities**
- **aligned entities**

**Supervised training**

- Induce new (highly confident) links
- Continue supervised training
- Induce more links

**Model**

- model 1.0
- model 1.1
- ...
Experimental Results
Two settings

- **Semi-supervised setting**
  - 4.5k ground truth label available
  - Iterative Learning is used to expand the label set

- **Unsupervised setting**
  - no ground truth label available
  - Iterative Learning is used to expand the label set
Semi-supervised setting: DBP15k

Comparing EVA (green) with the best baselines, Hits@1 is reported:

Without Iterative learning

With Iterative learning
Semi-supervised setting: DWY15k-DW

Comparing EVA (green) with the best baselines:

Without title surface forms

<table>
<thead>
<tr>
<th></th>
<th>normal</th>
<th>dense</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hits@1</td>
<td>45</td>
<td>67.5</td>
</tr>
</tbody>
</table>

With title surface forms

<table>
<thead>
<tr>
<th></th>
<th>normal</th>
<th>dense</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hits@1</td>
<td>22.5</td>
<td>45</td>
</tr>
</tbody>
</table>
Unsupervised setting

Comparing unsupervised performance with different #visual seeds:
(Results reported on DBP15k.)
Ablation study: which modality matters the most?

(Results are reported on DBP15k, FR->EN)
Qualitative analysis: long-tail entities benefit from images the most

Entity frequency in KGs follow long-tailed distribution.
As an example: we plot #appearances of 100 randomly sampled entities from DBP15k (FR-EN):
Qualitative analysis: long-tail entities benefit from images the most

\[ \text{DegSum}(e_s, e_t) := \text{deg}(e_s) + \text{deg}(e_t) \]

(Results are reported on DBP15k, FR->EN)
Thanks for watching!

Code & data:  https://github.com/cambridgeltl/eva
Contact:  fl399@cam.ac.uk

Acknowledgement:  several slides are borrowed from:
https://cogcomp.seas.upenn.edu/page/tutorial.202002/handout/3-mrrl.pdf