LNN-EL: A Neuro-Symbolic Approach for Short-text Entity Linking

Hang Jiang, Sairam Gurajada, Qiuhao Lu, Sumit Neelam, Lucian Popa, Prithviraj Sen, Yunyao Li, and Alexander Gray
Entity Linking on Short Text

Ex: “List all boardgames by GMT”

Why is it important:
• Better understanding of text
• Question-answering

Challenges
• Needs good exploitation of the context:
  o To match GMT with confidence, need available clues: co-occurring entities (e.g., “boardgames”), relationships, sentence, KB text and graph

http://lookup.dbpedia.org/api/search/KeywordSearch?QueryString=%22GMT%22&MaxHits=100

1 <Label>Greenwich Mean Time</Label>
<URI>http://dbpedia.org/resource/Greenwich_Mean_Time</URI>
<Description>Greenwich Mean Time (GMT) is a time system originally referring to mean solar time at the Royal Observatory in Greenwich, London, which later became adopted as a global time standard. It is arguably the same as Coordinated Universal Time (UTC) and when this is viewed as a time zone the name Greenwich Mean Time is especially used by bodies connected with the United Kingdom, such as the BBC World Service, the Royal Navy, the Met Office and others.</Description>
...

23 <Label>GMT Games</Label>
<URI>http://dbpedia.org/resource/GMT_Games</URI>
<Description>GMT Games, probably the most prolific of the wargame companies in the 1990s and 2000s, was founded in 1990. The current management and creative team includes Tony Curtis, Rodger MacGowan, Mark Simonitch, and Andy Lewis. The company has become well known for graphically attractive games that range from "monster games", of many maps and counters, to quite simple games suitable for introducing new players to wargaming. They also produce card games and family games.</Description>
Entity Linking Approach

**Mention Extraction**

Input text: “List all boardgames by GMT”

- **M1**: Boardgames
- **M2**: GMT

**DBpedia Lookup API**

- Mentions (S)
- Candidate List

**Rule-based Disambiguation**

For mention: M1

- <Label>GMT</Label>
- <URI>dbpedia.org/resource/GMT</URI>
- ...

For mention: M2

- <Label>GMT Games</Label>
- <URI>dbpedia.org/resource/GMT_Games</URI>
- <Description>...well known for graphically attractive games that range from "monster games..."
Rule-based Entity Disambiguation with LNN-EL

**User provided EL Algorithm**

\[ R_1(m_i, e_{ij}) \leftarrow f_1(m_i, e_{ij}) > \theta_1 \land f_2(m_i, e_{ij}) > \theta_2 \land f_3(m_i, e_{ij}) > \theta_3 \]

\[ R_2(m_i, e_{ij}) \leftarrow f_1(m_i, e_{ij}) > \theta_4 \land f_4(m_i, e_{ij}) > \theta_5 \]

**Symbolic Rules**

**Extensible space of features:**

**Non-embedding based**

- String similarity (Jaccard, JaroWinkler, Levenshtein, etc.)
- Candidate importance score (e.g., refCount)
- Type similarity
- ...

**Embedding based**

- BERT, Wiki2Vec
- Query2Box embeddings,
- Scores of prior EL methods (e.g., BLINK)

**Neural Learning**

**Learnable parameters:**

- \( \theta_i \): feature thresholds,
- \( f_{w_i} \): feature weights,
- \( r_{w_i} \): rule weights

Based on a learnable real-valued logic framework (LNN):

[R. Riegel et al. *Logical Neural Networks*, 2020]
Rule-based Entity Disambiguation with LNN-EL

.User provided EL Algorithm
\[
R_1(m_i, e_{ij}) \leftarrow f_1(m_i, e_{ij}) > \theta_1 \land f_2(m_i, e_{ij}) > \theta_2 \\
\land f_3(m_i, e_{ij}) > \theta_3 \\
\lor \\
R_2(m_i, e_{ij}) \leftarrow f_1(m_i, e_{ij}) > \theta_4 \land f_4(m_i, e_{ij}) > \theta_5
\]

Symbolic Rules

Advantages of LNN-EL
1. interpretable: expressive FOL language
2. extensible
3. transferable

Based on a learnable real-valued logic framework (LNN):
[R. Riegel et al. Logical Neural Networks, 2020]
Entity Linking Performance on Benchmark Datasets

<table>
<thead>
<tr>
<th>Model</th>
<th>LC-QuAD</th>
<th>QALD-9</th>
<th>WebQSP_EL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Precision</td>
<td>Recall</td>
<td>F1</td>
</tr>
<tr>
<td>BLINK</td>
<td>87.04</td>
<td>87.04</td>
<td>87.04</td>
</tr>
<tr>
<td>BERT</td>
<td>57.14</td>
<td>63.09</td>
<td>59.97</td>
</tr>
<tr>
<td>BERTWiki</td>
<td>66.96</td>
<td>73.85</td>
<td>70.23</td>
</tr>
<tr>
<td>Box</td>
<td>67.31</td>
<td>74.32</td>
<td>70.64</td>
</tr>
<tr>
<td>LogisticRegression</td>
<td>87.04</td>
<td>86.83</td>
<td>86.93</td>
</tr>
<tr>
<td>LogisticRegression_{BLINK}</td>
<td>90.50</td>
<td>90.30</td>
<td>90.40</td>
</tr>
</tbody>
</table>

### Baselines
- RuleEL
- LogicEL
- LNN-EL
- LNN-EL_{_{\text{_{\text{_{BLINK}}}}}}

### Logic-based (ours)

**LNN-EL:**
- Reaches SotA for entity linking on KBQA datasets
  - Improves on BLINK [Wu et al, 2020], a black-box zero-shot model based on BERT, pre-trained on 9M Wikipedia examples
  - Easily extensible
    - Ensembles that are progressively richer in features: string similarity, BERT embeddings, Box embeddings, BLINK
- Interpretability
Extensibility: A Closer Look at the Rules

**Name Rule:**

\[
R_{\text{name}} \leftarrow \left[ f_{\text{acc}}(m_i, e_{ij}) > \theta_1 \lor f_{\text{lev}}(m_i, e_{ij}) > \theta_2 \lor f_{\text{wv}}(m_i, e_{ij}) > \theta_3 \lor f_{\text{spacy}}(m_i, e_{ij}) > \theta_4 \right] \land f_{\text{prom}}(m_i, e_{ij})
\]

Name similarity

**Context Rule:**

\[
R_{\text{ctx}} \leftarrow \left[ f_{\text{acc}}(m_i, e_{ij}) > \theta_1 \lor f_{\text{lev}}(m_i, e_{ij}) > \theta_2 \lor f_{\text{wv}}(m_i, e_{ij}) > \theta_3 \lor f_{\text{spacy}}(m_i, e_{ij}) > \theta_4 \right] \land f_{\text{ctx}}(m_i, e_{ij}) \land f_{\text{prom}}(m_i, e_{ij})
\]

Context similarity of co-mentions and DBpedia entity desc.

**Type Rule:**

\[
R_{\text{type}} \leftarrow \left[ f_{\text{acc}}(m_i, e_{ij}) > \theta_1 \lor f_{\text{lev}}(m_i, e_{ij}) > \theta_2 \lor f_{\text{wv}}(m_i, e_{ij}) > \theta_3 \lor f_{\text{spacy}}(m_i, e_{ij}) > \theta_4 \right] \land f_{\text{type}}(m_i, e_{ij}) \land f_{\text{prom}}(m_i, e_{ij})
\]

Type similarity

**Blink Rule:**

\[
R_{\text{blink}} \leftarrow \left[ f_{\text{acc}}(m_i, e_{ij}) > \theta_1 \lor f_{\text{lev}}(m_i, e_{ij}) > \theta_2 \lor f_{\text{wv}}(m_i, e_{ij}) > \theta_3 \lor f_{\text{spacy}}(m_i, e_{ij}) > \theta_4 \right] \land f_{\text{blink}}(m_i, e_{ij})
\]

BLINK score

**Box Rule:**

\[
R_{\text{box}} \leftarrow \left[ f_{\text{acc}}(m_i, e_{ij}) > \theta_1 \lor f_{\text{lev}}(m_i, e_{ij}) > \theta_2 \lor f_{\text{wv}}(m_i, e_{ij}) > \theta_3 \lor f_{\text{spacy}}(m_i, e_{ij}) > \theta_4 \right] \land f_{\text{box}}(m_i, e_{ij}) > \theta_5
\]

Box similarity (co-mentions and DBpedia neighboring nodes)

---

**With more features available:**
- Performance typically increases.
- Combining the features into rules also becomes more challenging (**full-fledged rule learning will be needed**).
Transferability

• Inductive bias offered by using rules leads to good transfer across different datasets within the same domain.

• No fine-tuning on the target dataset

• LNN-EL performs reasonably well, even in cases where training is done on a very small dataset.
  – E.g., from QALD-9 (with only a few hundred questions to train) to WebQSP: F1-score of 83.06 (vs. 85.08)

• Our competitor, zero-shot BLINK by design has very good transferability too, but it is trained on entire Wikipedia.

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Test</th>
<th>WebQSP_EL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LC-QuAD</td>
<td>QALD-9</td>
<td></td>
</tr>
<tr>
<td>LC-QuAD</td>
<td>87.64</td>
<td>86.41</td>
<td>78.90</td>
</tr>
<tr>
<td>QALD-9</td>
<td>85.58</td>
<td>88.52</td>
<td>83.06</td>
</tr>
<tr>
<td>WebQSP_EL</td>
<td>80.95</td>
<td>87.25</td>
<td>85.08</td>
</tr>
</tbody>
</table>
Summary & Future Directions

• Summary
  • LNN-EL, a neuro-symbolic approach for entity linking on short text
  • Achieved competitive performance against SotA black-box neural models
  • LNN-EL is **interpretable**, **extensible** and **customizable**
  • LNN-EL may **transfer** better to new datasets

• Future directions
  – Automatic learning of the rule templates
  – Longer documents
Thank you!