Entity Enabled Relation Linking

Jeff Z. Pan, Mei Zhang, Kuldeep Singh, Frank Van Harmelen, Jinguang Gu, Zhi Zhang
Relation Linking

• The task of **Relation Linking (RL)**
  – Given some NL text q, and a KG G, identify relevant relations in G
  – Example:
    • q: Where was Cliff Richard born?
    • G: DBpedia => dbo:birthPlace
  – Motivations
    • **KG building:** q as a document, such as a news article in Fake News Detection [ISWC2018]
    • **Question understanding and answering:** q as a question

• The task of **Semantic Parsing (SP)**
  – Transform some NL question into a formal query against a database or a **knowledge base**
  – Related tasks: Entity Linking (EL) and Query Building
Research Questions

• The task of Semantic Parsing (SP)
  – Transform some NL question into a formal query against a database or a knowledge base
  – Related tasks: Entity Linking (EL) and Query Building

• RQ1: how to take advantage of the knowledge graph for relation linking?
• RQ2: should EL and RL be run sequentially, parallelly or jointly?
Research Questions

• **RQ1**: how to take advantage of the knowledge graph for relation linking?
  – What are the potential benefits of using KG
    • KG as a global schema of multiple databases sharing relations
    • Fine-grained semantics: e.g. 2019 (Year), 200km
    • Connections (such as temporal connections) among entities; such as Olympic 2000, 2004, 2008, ...
  – Existing work
    • Keyword based relation expansion with external resources and similarity measures
      – **Rematch**: using Wordnet
      – **SIBKB**: using PATTY
Research Questions

• **RQ1**: how to take advantage of the knowledge base for relation linking?
  – Existing work
    • **EARL**: Transforming KG into subdivision graphs for joint entity and relation linking

• **RQ2**: should EL and RL be run sequentially, parallelly or jointly?
  – Sequentially
  – Parallelly
  – jointly. (EARL)
Research Questions

• **RQ1’**: Is there any other ways of taking advantage of the knowledge base for relation linking?

• **RQ2’**: Would it be beneficial if EL is performed before RL?
Problem Statement

- **Problem**: Entity Enabled Relation Linking
  - Given input question q and a knowledge graph G=S U D, and set of Entities E_q ⊆ E(D) identified in q
  - the task is to identify set of relations R_q ⊆ R(S), for the set of relation phrases in q based on the entities E_q.
Example: Relation Linking over DBpedia

• **Q1**: Which comic characters are painted by Bill Finger? (dbr:Bill_Finger)
  – Answer: dbo:paint?
  – Answer: dbo:painter?
Example: Relation Linking over DBpedia

- **Q2**: Where was Cliff Richard born? (dbr:Cliff_Richard)
  - Answer: dbo:birthPlace

About: Cliff Richard

Sir Cliff Richard OBE (born Harry Rodger Webb, 14 October 1940) is a British pop singer, musician, performer, actor, and philanthropist. Richard has sold more than 250 million records worldwide. He has total sales of over 21 million singles in the United Kingdom and is the third-top-selling artist in UK Singles Chart history, behind the Beatles and Elvis Presley. Richard has been a resident in the United Kingdom for most of his life, but in 2010, he confirmed that he had become a citizen of Barbados. He divides his time between living in Barbados and Portugal.
Example: Relation Linking over DBpedia

• **Q2**: Where was Cliff Richard born? (db:Cliff_Richard)
  – Answer: dbo:birthPlace

• **Q3**: Where was Richard Archer born? (db:Richard_Archer)
  – Answer: dbo:birthPlace?
Example: Relation Linking over DBpedia

• **Q2**: Where was Cliff Richard born? (dbr:Cliff_Richard)
  – Answer: dbp:birthPlace?

• **Q3**: Where was Richard Archer born?
  (dbr:Richard_Archer)
  – Answer: dbo:birthPlace?
  – Answer: dbo:hometown
Observations

• **O1**: Cannot purely rely on similarity measures
• **O2**: Need to query the graph
  – data sub-graph
  – schema sub-graph
Entity based Relation Expansion

- **Explicit relations**: direct relations of $E_q$ in $G=<S,D>$
  - $\{R | <E_q, R, _> \in D\} \cup \{R | <_, R, E_q> \in D\}$

- **Implicit relations**: relations in $G$ with global/local domain/range as the types of $E_q$
  - $\{R | <E_q, \text{rdf:type } GD> \in D \text{ and } S |= \exists R . T \sqsubseteq GD\} \cup$
  - $\{R | <E_q, \text{rdf:type } LD> \in D \text{ and } S |= \exists R . C \sqsubseteq LD\} \cup$
  - $\{R | <E_q, \text{rdf:type } GR> \in D \text{ and } S |= \exists R^- . T \sqsubseteq GR\} \cup$
  - $\{R | <E_q, \text{rdf:type } LR> \in D \text{ and } S |= \exists R^- . C \sqsubseteq LR\}$
Example Revisited: Bill Finger

- **Q1:** Which comic characters are painted by Bill Finger? (dbr:Bill_Finger)
  - Answer: dbo:paint?
  - Answer: dbo:painter?
  - dbr:Bill_Finger has type dbo:Agent
  - dbo:creator has range dbo:Agent
  - dbr:Bill_Finger has relation dbo:creator
Example Revisited: Ricard Archer

• **Q3**: Where was Richard Archer born? (dbr:Richard_Archer)
  – dbr:Richard_Archer has type dbo:Person
  – dbo:birthplace has range dbo:Person
  – However, dbr:Richard_Archer does **not** have relation **dbo:birthplace**
  – dbr:Richard_Archer has type dbo:Agent
  – **dbo:hometown** has range dbo:Agent
  – dbr:Richard_Archer has relation **dob:hometown**
Experimental Evaluation

• Datasets:
  – Question answering over Linked Data challenge (QALD): 58% with simple questions
    • QALD-5 have 340 questions, and QALD-7 has 215 questions
  – LC-QuAD: 5000 questions, with 20% simple questions

• SOTA systems
  – SIBKB
  – Rematch
  – EARL

• Experimental settings
  – 1 virtual server, with eight cores and 32 GB RAM running on the Ubuntu 16.04.3 operating system
  – Frankenstein Resource Platform
Results

- Our system outperforms baseline systems in all 3 datasets
- Unlike the baseline systems, ours does not have major performance drop for more complex questions

<table>
<thead>
<tr>
<th>QA Component</th>
<th>Dataset</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIBKB</td>
<td>QALD-5</td>
<td>0.27</td>
<td>0.34</td>
<td>0.29</td>
</tr>
<tr>
<td>ReMatch</td>
<td>QALD-5</td>
<td>0.36</td>
<td>0.39</td>
<td>0.37</td>
</tr>
<tr>
<td>EARL</td>
<td>QALD-5</td>
<td>0.17</td>
<td>0.21</td>
<td>0.19</td>
</tr>
<tr>
<td>EERL</td>
<td>QALD-5</td>
<td>0.43</td>
<td>0.49</td>
<td>0.45</td>
</tr>
<tr>
<td>SIBKB</td>
<td>QALD-7</td>
<td>0.33</td>
<td>0.35</td>
<td>0.34</td>
</tr>
<tr>
<td>ReMatch</td>
<td>QALD-7</td>
<td>0.35</td>
<td>0.38</td>
<td>0.37</td>
</tr>
<tr>
<td>EARL</td>
<td>QALD-7</td>
<td>0.30</td>
<td>0.31</td>
<td>0.30</td>
</tr>
<tr>
<td>EERL</td>
<td>QALD-7</td>
<td>0.42</td>
<td>0.46</td>
<td>0.43</td>
</tr>
<tr>
<td>SIBKB</td>
<td>LC-QuAD</td>
<td>0.15</td>
<td>0.18</td>
<td>0.16</td>
</tr>
<tr>
<td>ReMatch</td>
<td>LC-QuAD</td>
<td>0.18</td>
<td>0.20</td>
<td>0.19</td>
</tr>
<tr>
<td>EARL</td>
<td>LC-QuAD</td>
<td>0.20</td>
<td>0.25</td>
<td>0.21</td>
</tr>
<tr>
<td>EERL</td>
<td>LC-QuAD</td>
<td>0.53</td>
<td>0.58</td>
<td>0.55</td>
</tr>
</tbody>
</table>
Conclusion & Future Work

- **RQ1**: how to take advantage of the knowledge graph for relation linking?
  - Structure in both data and schema sub-graphs can help
  - Reasoning can be performed offline

- **RQ2**: should EL and RL be run sequentially, parallelly or jointly?
  - Entity Linking can help Relation Linking

- **Future work**
  - Alternative ranking methods
  - Deal with incomplete schema
  - Potential new ways of coordinating EL and RL
  - Classify hard questions
Thank you!

jeff.z.pan@abdn.ac.uk
http://knowledge-representation.org/j.z.pan/
@jpansw