When a Knowledge Base is not Enough

Question Answering over Knowledge Bases with External Text Data

Denis Savenkov
Emory University
dsavenk@emory.edu

Eugene Agichtein
Emory University
eugene@mathcs.emory.edu

SIGIR 2016
Percentage of question search queries is growing\[1\]

\[1\] “Questions vs. Queries in Informational Search Tasks”, Ryen W. White et al, WWW 2015
Automatic Question Answering works relatively well for simple factoid questions.
For many questions we still have to dig into “10 blue links”

* the questions are taken from different QA datasets (WebQuestions, QALD-5, Yahoo! Answers Webscope)
Different data sources are used for question answering

<table>
<thead>
<tr>
<th>Unstructured data</th>
<th>Semi-structured data</th>
<th>Structured data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text documents</td>
<td>Web tables &amp; infoboxes</td>
<td>Knowledge bases</td>
</tr>
</tbody>
</table>
Data Sources have different advantages and problems

**Text documents**
- easy to match against question text
- cover a variety of different information types
- each text phrase encodes a limited amount of information about mentioned entities

**Knowledge bases**
- aggregate the information around entities
- allow complex queries over this data using special languages (e.g. SPARQL)
- hard to translate natural language questions into special query languages
- incomplete (missing entities, facts and properties)
Advantages of one Data Source can compensate disadvantages of the other

Text documents

- easy to match against question text
- cover a variety of different information types
- each text phrase encodes a limited amount of information about mentioned entities

Knowledge bases

- hard to translate natural language questions into special query languages
- incomplete (missing entities, facts and properties)
- aggregate the information around entities
Knowledge Base Question Answering (KBQA)

- **Goal**: translate natural language question into structured KB query (e.g. SPARQL) to retrieve correct entity or attribute value

**When did Tom Hanks win his first Oscar?**

```sparql
PREFIX fb: <http://rdf.freebase.com/ns/>
SELECT ?year WHERE {
  fb:/m/0bxtg fb:/award/award_winner/awards_won ?award .
  ?award fb:/award/award_honor/award fb:/m/0f4x7 .
  ?nomination fb:/award/award_honor/year ?year .
} ORDER BY ?year LIMIT 1
```
Knowledge Base Question Answering Challenges

1. Query analysis
   ○ How to identify question topic entity to anchor KB search?

2. Candidate generation
   ○ What predicates might correspond to words and phrases in the question?
   ○ What entities to include as candidate answers?

3. Evidence extraction
   ○ How to score correspondence between a certain candidate answer (e.g. involved predicates) and the question?

4. Answer selection
   ○ How to rank candidate answers to select the final response?
Existing Text-KB hybrid approaches

✓ **Open QA** [A.Fader et al. 2014]
  → Use Open Information Extraction to build semi-structured KB from text
  → Joint QA over extracted and curated KB

✓ **Extended Knowledge Graphs** [S. Elbassuoni et al 2009, M.Yahya et al 2016]
  → Extend triples in knowledge base with keywords
  → SPARQL query relaxation techniques to use keyword matches

✓ **“Open Domain Question Answering via Semantic Enrichment”** [H.Sun et al 2015]
  → Annotate text with entity mentions
  → Use entity types and textual KB descriptions to improve text-based QA

✓ **“Question Answering on Freebase via Relation Extraction and Textual Evidence”** [K. Xu et al. 2016]
  → Using text documents to refine answers, generated by KBQA system

✓ **Memory Networks** [A. Bordes et al 2015]
  → Encode curated and OpenIE triples into NN memory
Text2KB: main idea

✓ Improve different stages in Knowledge Base Question Answering using various textual data
  ○ query analysis
    ✓ question topic entity identification using web search results
  ○ candidate generation
    ✓ Mine associations patterns between question terms and predicates from CQA data
  ○ evidence extraction
    ✓ build language model for candidate question-answer entity pairs based on annotated corpus of text documents
  ○ answer selection
    ✓ Score answer candidates using a combination of KB and text-based features
Text2KB: Incorporating Text in Answering Process
Baseline system architecture*

1. **Detecting question topic entity:** multiple candidates are detected using dictionary of names and aliases

2. **Answer candidate generation:** instantiate candidate SPARQL queries from the neighborhood of question entities using a set of template queries

3. **Evidence generation:** each candidate is represented with a set of features, describing the detected topic entity, predicates on KB path connecting topic and answer entities, etc.

4. **Answer selection:** candidate answers are ranked using a trained ranking model and top scoring one is returned as the answer

* “More Accurate Question Answering on Freebase” by Hannah Bast et al, 2015
Text2KB System Architecture

Existing KBQA system

Text-based resources to improve KBQA
Question Analysis: Entity Linking

✓ Web Search Results can help entity linking and provide textual evidence to answer candidates
✓ Contains multiple mentions of the question topic entity, often in variations, which might help entity linking
✓ Search results often contain the answer to the question itself, which is exploited by text-based question answering systems
Text2KB System Architecture: web search results

- Top 10 results using Bing Web Search API & Wikipedia Search
- Identify mentioned KB entities using QA system’s entity linking module
  - Extend the set of question topic entity
  - Use mention counts as features for candidate ranking
Community Question Answering data can help map question phrases to predicates

✓ Huge number of question-answer pairs, but noisy (most of the questions aren’t factoid, answers are verbose and contain redundant information)

✓ Can be helpful to learn associations between the language of a question and KB predicates using distant supervision assumption
Examples of term-predicate associations computed using CQA data

<table>
<thead>
<tr>
<th>Term</th>
<th>Predicate</th>
<th>PMI score</th>
</tr>
</thead>
<tbody>
<tr>
<td>born</td>
<td>people.person.date_of_birth</td>
<td>3.67</td>
</tr>
<tr>
<td></td>
<td>people.person.date_of_death</td>
<td>2.73</td>
</tr>
<tr>
<td></td>
<td>location.location.people_born_here</td>
<td>1.60</td>
</tr>
<tr>
<td>kill</td>
<td>people.deceased_person.cause_of_death</td>
<td>1.70</td>
</tr>
<tr>
<td></td>
<td>book.book.characters</td>
<td>1.55</td>
</tr>
<tr>
<td>currency</td>
<td>location.country.currency_formerly_used</td>
<td>5.55</td>
</tr>
<tr>
<td></td>
<td>location.country.currency_used</td>
<td>3.54</td>
</tr>
<tr>
<td>school</td>
<td>education.school.school_district</td>
<td>4.14</td>
</tr>
<tr>
<td></td>
<td>people.education.institution</td>
<td>1.70</td>
</tr>
<tr>
<td></td>
<td>sports.school_sports_team.school</td>
<td>1.69</td>
</tr>
<tr>
<td>win</td>
<td>sports.sports_team.championships</td>
<td>4.11</td>
</tr>
<tr>
<td></td>
<td>sports.sports_league.championship</td>
<td>3.79</td>
</tr>
</tbody>
</table>

✓ Despite the noisy distant supervision labeling, top scoring predicates are indeed related to the corresponding word
Text2KB System Architecture: CQA data

- Distant supervision to label question-answer pairs from Yahoo! Answers WebScope collection with KB predicates
- Learn associations between question terms and predicates using PMI scores
  - Use these PMI scores as features to score candidate answer predicates
Text around mentions of pairs of entities in documents help explain relationships between the entities

✓ Sentences and passages that mention multiple entities often express some facts about them
✓ Terms used in these passages can explain the relationships between the entities
Examples of entity pair language models

<table>
<thead>
<tr>
<th>Entity 1</th>
<th>Entity 2</th>
<th>Term counts</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Edwards</td>
<td>Rielle Hunter</td>
<td>campaign, affair, mistress, child, former...</td>
</tr>
<tr>
<td>John Edwards</td>
<td>Cate Edwards</td>
<td>daughter, former, senator, courthouse, greensboro, eldest...</td>
</tr>
<tr>
<td>John Edwards</td>
<td>Elizabeth Edwards</td>
<td>wife, hunter, campaign, affair, cancer, rielle, husband...</td>
</tr>
<tr>
<td>John Edwards</td>
<td>Frances Quinn</td>
<td>daughter, john, rielle, father, child, former, paternity...</td>
</tr>
</tbody>
</table>

✓ Terms most frequently used around mention of a pair of entities indeed shed some light on the relationship between the entities
Text2KB System Architecture: document collection

- Extract text around mentions of entity pairs in ClueWeb12
- Learn entity pair language model $p(\text{term} | \text{entity}_1, \text{entity}_2)$
  - Use language model scores as features for candidate answer ranking
Evaluation

✓ WebQuestions dataset
  ○ 3,778 training and 2,032 test questions
✓ Metrics:
  ○ Average F1:
  \[
  \text{avg } F1 = \frac{1}{|Q|} \sum_{q \in Q} f1(a^*_q, a_q)
  \]
  \[
  f1(a^*_q, a_q) = 2 \frac{\text{precision}(a^*_q, a_q) \text{recall}(a^*_q, a_q)}{\text{precision}(a^*_q, a_q) + \text{recall}(a^*_q, a_q)}
  \]
✓ Methods compared:
  ○ Aqqu (Bast et al, 2015) - our KB-only baseline
  ○ STAGG (Yih et al, 2015) - SOTA at the moment of publication
  ○ our Text2KB (Web search)
  ○ our Text2KB (Wikipedia search)
Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Recall</th>
<th>Precision</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>OpenQA [A.Fader et al 2014]</td>
<td>-</td>
<td>-</td>
<td>0.35</td>
</tr>
<tr>
<td>STAGG [H.Sun et al 2015]</td>
<td>0.607</td>
<td>0.528</td>
<td>0.525</td>
</tr>
<tr>
<td>Aqqu (baseline) [H.Bast et al 2015]</td>
<td>0.604</td>
<td></td>
<td>0.494</td>
</tr>
<tr>
<td>Text2KB (wikipedia search)</td>
<td>0.632</td>
<td>0.498</td>
<td>0.514</td>
</tr>
<tr>
<td>Text2KB (web search)</td>
<td>0.635</td>
<td>0.506</td>
<td>0.522</td>
</tr>
</tbody>
</table>

✓ Text2KB significantly improves upon the baseline Aqqu system (0.494 -> 0.522 avg F1 score)

✓ Text2KB reaches the performance of STAGG, best result at the moment of publication
  ○ but this work is orthogonal to improvements in STAGG and therefore can be combined
Component ablation

<table>
<thead>
<tr>
<th>System</th>
<th>avg F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aqqu</td>
<td>0.494</td>
</tr>
<tr>
<td>+ Entity linking from search results</td>
<td>0.508</td>
</tr>
<tr>
<td>+ Search results, CQA and ClueWeb features for ranking</td>
<td>0.514</td>
</tr>
<tr>
<td>Text2KB</td>
<td>0.522</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>System</th>
<th>avg F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aqqu</td>
<td>0.494</td>
</tr>
<tr>
<td>Text2KB (Web search)</td>
<td>0.522</td>
</tr>
<tr>
<td>- Web search data</td>
<td>0.513</td>
</tr>
<tr>
<td>- CQA data</td>
<td>0.519</td>
</tr>
<tr>
<td>- ClueWeb data</td>
<td>0.523</td>
</tr>
<tr>
<td>+ Web search data only</td>
<td>0.522</td>
</tr>
<tr>
<td>+ CQA data only</td>
<td>0.508</td>
</tr>
<tr>
<td>+ ClueWeb data only</td>
<td>0.514</td>
</tr>
</tbody>
</table>

 ✓ Both entity linking using web search results and features for answer ranking contribute to improvements
 ✓ Search results have the largest contribution to the overall performance, but CQA and ClueWeb are also useful
Combining Text2KB & STAGG

<table>
<thead>
<tr>
<th>System</th>
<th>avg F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAGG (Yih et al, 2015)</td>
<td>0.525</td>
</tr>
<tr>
<td>Text2KB + STAGG (takes STAGG answers if it has less entities)</td>
<td>0.532</td>
</tr>
<tr>
<td>Text2KB + STAGG (Oracle: chooses answer with higher F1 score)</td>
<td>0.606</td>
</tr>
</tbody>
</table>

✓ Combining results of Text2KB and STAGG suggests that our ideas could benefit it as well
  ○ Heuristic combination: take Text2KB or STAGG answer, which contains less entities
  ○ Oracle combination always choose the answer with higher F1
Error analysis

✓ Majority of errors (F1 < 1) are ranking errors
✓ But there are also many problems in questions and labels
✓ Check out the new WebQuestionsSP dataset: https://goo.gl/eQF0tM
Current & Future work

○ Overall, our system is most helpful:
  ➢ Question topic entity is hard to identify (uncommon alias, misspelling)
  ➢ Form of the question or ground truth predicate is less frequent in the training set

○ Our system has the following problems:
  ➢ Less effective for tail and abstract entities, whose mentions are harder to find in text. For example entity “Associated Press Male Athlete of the Year” isn't linked correctly (unless mentioned exactly by name)
  ➢ Our use of text doesn't help much to solve KB incompleteness (e.g. missing facts or predicates)

○ Future work:
  ➢ Instead of improving KBQA, move to more open scenario
    ■ new hybrid model that will use all the information available in different data sources
    ■ new dataset of entity-centric factoid questions
Conclusions

- Textual data sources provide additional information, that can compensate disadvantages of structured knowledge bases.
- Our Text2KB system uses a combination of structured and unstructured data to improve Knowledge Base Question Answering.
  - Improve avg F1 on WebQuestions dataset: 0.494 -> 0.522
Acknowledgements

Denis Savenkov is planning to defend in December 2016 and will be on the market for postdoc and industry research positions

dsavenk@emory.edu

Thank you!

This work was partially supported by the Yahoo Labs Faculty Research Engagement Program (FREP).