Feature Driven Question Answering

Benjamin Van Durme
Question Answering (QA)

Natural Language Interface to Databases

Information Retrieval (+ Information Extraction)
Question Answering (QA)

Natural Language Interface to Databases

Information Retrieval (+ Information Extraction)
1960’s, 1970’s, …

build semantic parsers by hand
Baseball (Green et al., 1961)

“Where did the Red Sox play on July 7?”

[Where] did [the Red Sox] play (on [July 7])?

[ANSWER]
Lunar (Woods, 1977)

“What is the concentration of silicon in S10046?”

(For The X / (Dataline S10046 Overall SIO2) : T ;
(For Every Y / Thing : (Equal X Y) ;
(Printout Y) ))) → [ANSWER]
[fast forward]
Difference?

1960’s, 1970’s, …
   build grammars by hand

1990’s, 2000’s
   induce grammars from annotations
Zelle and Mooney (1996)

“What is the capital of the state with the largest population?”

answer(C,(capital(S,C),
    largest(P, (state(S), population(S,P))))))
Zettlemoyer and Collins (2005)

“What states border texas?”

$\lambda x.\text{state}(x) \land \text{borders}(x, \text{texas})$
1960’s, 1970’s, ...
    build grammars by hand
1990’s, 2000’s
    induce grammars from annotated questions
2010’s
    induce grammars based on questions + answers
Clarke et al. (2010)

“What is the largest state that borders Texas?”

New Mexico
Clarke et al. (2010)

“What is the largest state that borders Texas?”

\[ \text{largest(state(next\_to(const(texas))))} \]
Structured Query

Parse

Logic

Structured Query

Result
Question Answering (QA)

Natural Language Interface to Databases

Information Retrieval (+ Information Extraction)
Question Answering (QA)

Natural Language Interface to Databases

Information Retrieval (+ Information Extraction)
“What is the something something something?”

\[ \text{Lambda } x \ [\text{constraint1}(x), \text{constraint2}(x), \ldots] \]
“What is the something something something?”

\[ \text{Lambda } x \ [\text{constraint1}(x), \text{constraint2}(x), ...] \]
"What is the something something something?"

Lambda x [constraint1(x), constraint2(x), ...]
“What is the something something something?”

Lambda x [constraint1(x), constraint2(x), ...]
“What is the something something something?”
NIST TREC-8 QA
(Voorhees 1999)

200 questions, each with “snippet” answers

“How many calories in a Big Mac?”

There are 549 calories in 1 burger (7.6 oz) of McDonald's Big Mac Burger. You'd need to walk …
Chen and Van Durme (in-progress)
Yao and Van Durme (2014)
Yao et al. (2014)
Yao et al. (2013a,b,c)
Van Durme et al. (2003)
Nyberg et al. (2003)
Chen and Van Durme (in-progress)
Yao and Van Durme (2014)
Yao et al. (2014)
Yao et al. (2013a,b,c)
Van Durme et al. (2003)
Nyberg et al. (2003)
Question Answering (QA)

Information Retrieval (+ Information Extraction)
- Feature driven answer extraction
- Feature driven retrieval
- Feature driven QA on a KB

Johns Hopkins University
Feature driven answer extraction
Feature driven retrieval
Feature driven QA on a KB
“What sport does Jennifer Capriati play?”
“What sport does Jennifer Capriati play?”

“Capriati hasn't played on the tour since ...”
“Tennis player Jennifer Capriati is 23”
“Jennifer Capriati enjoys sports”
...
“What sport does Jennifer Capriati play?”

“Capriati hasn't played on the tour since ...”

“Tennis player Jennifer Capriati is 23”

“Jennifer Capriati enjoys sports”

...
“What sport does Jennifer Capriati play?”

“Capriati hasn’t played on the tour since …”

“Tennis player Jennifer Capriati is 23”

“Jennifer Capriati enjoys sports”

…
“What sport does Jennifer Capriati play?”

Tennis player Jennifer Capriati is 23
“What sport does Jennifer Capriati play?”

“BIO” tagging, 3 types of hidden states:
- B-ANS (beginning of answer)
- I-ANS (inside of answer)
- O (outside of answer, i.e., not an answer)
Answer Extraction as Sequence Tagging
using a linear-chain Conditional Random Field (CRF)

$$p(y | x) = \frac{1}{Z(x)} \prod_{t=1}^{T} \exp \left\{ \sum_{k=1}^{K} \theta_k f_k (y_t, y_{t-1}, x_t) \right\}$$

hidden states $Y$

observation $X$

conditional model $p(y|x)$:

$f(y_t, y_{t-1}, x_t)$: feature functions

$\theta$: feature weights (to learn)
Answer Extraction as Sequence Tagging using a linear-chain Conditional Random Field (CRF)

conditional model $p(y|x)$:

\[ f(y_t, y_{t-1}, x_t) : \text{feature functions} \]
\[ \theta : \text{feature weights (to learn)} \]
NLP is full of features

does question starts with “what”, “where”, “who”, … ?
what parts of speech?
any named entities, like a PERSON or a LOCATION ?
automatic alignments between question and passage
(tree) edit distance between question and passage
…
Joint Features

Question starts with (“who”)
AND Passage contains entity of type (PERSON)

Question contains word (“born”)
AND Question starts with (“where”)
AND Passage contains word (“birthplace”)
Ranking Sentence Candidates

<table>
<thead>
<tr>
<th>System</th>
<th>MAP</th>
<th>MRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wang et al. (2007)</td>
<td>0.6029</td>
<td>0.6852</td>
</tr>
<tr>
<td>Heilman and Smith (2010)</td>
<td>0.6091</td>
<td>0.6917</td>
</tr>
<tr>
<td>Wang and Manning (2010)</td>
<td>0.5951</td>
<td>0.6951</td>
</tr>
<tr>
<td>this paper (48 features)</td>
<td>0.6319</td>
<td>0.7270</td>
</tr>
<tr>
<td>+WNsearch</td>
<td><strong>0.6371</strong></td>
<td><strong>0.7301</strong></td>
</tr>
<tr>
<td>+WNfeature (11 more feat.)</td>
<td>0.6307</td>
<td><strong>0.7477</strong></td>
</tr>
</tbody>
</table>
Feature driven answer extraction
Feature driven retrieval
Feature driven QA on a KB
Feature driven answer extraction
Feature driven retrieval
Feature driven QA on a KB
“When was Alaska purchased?”
“When was Alaska purchased?”
“When was Alaska purchased?”
“When was Alaska purchased?”

Standard search query: (alaska, purchased)
“When was Alaska purchased?”

“Eventually Alaska Airlines will allow all travelers who have purchased electronic tickets through any means …”
## Expected Answer Type

<table>
<thead>
<tr>
<th>Work</th>
<th>Question/Answer Types</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harabagiu et al. (2000)</td>
<td>27 named entity labels with many-to-many mapping to answer types of 15 top nodes</td>
<td>TREC</td>
</tr>
<tr>
<td>Hovy et al. (2000, 2002)</td>
<td>topology of 94 nodes (47 leaf nodes) analyzing 17,384 questions</td>
<td>answers.com</td>
</tr>
<tr>
<td>Hermjakob (2001)</td>
<td>122 question targets from above questions, taxonomy connecting 153 WordNet sub-hierarchies</td>
<td>answers.com</td>
</tr>
<tr>
<td>Pasca and Harabagiu (2001)</td>
<td></td>
<td>TREC</td>
</tr>
<tr>
<td>Li and Roth (2002)</td>
<td>6 coarse classes and 50 fine classes</td>
<td>TREC</td>
</tr>
<tr>
<td>Prager et al. (2006)</td>
<td>flat hierarchy with 80 semantic classes</td>
<td>TREC</td>
</tr>
</tbody>
</table>
## Expected Answer Type

<table>
<thead>
<tr>
<th>work</th>
<th>question/answer types</th>
<th>domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harabagiu et al. (2000)</td>
<td>27 named entity labels with many-to-many mapping to answer types of 15 top nodes</td>
<td>TREC</td>
</tr>
<tr>
<td>Hovy et al. (2000, 2002)</td>
<td>topology of 94 nodes (47 leaf nodes) analyzing 17,384 questions</td>
<td>answers.com</td>
</tr>
<tr>
<td>Hermjakob (2001)</td>
<td>122 question targets from above questions</td>
<td>answers.com</td>
</tr>
<tr>
<td>Pasca and Harabagiu (2001)</td>
<td>taxonomy connecting 153 WordNet sub-hierarchies</td>
<td>TREC</td>
</tr>
<tr>
<td>Li and Roth (2002)</td>
<td>6 coarse classes and 50 fine classes</td>
<td>TREC</td>
</tr>
<tr>
<td>Prager et al. (2006)</td>
<td>flat hierarchy with 80 semantic classes</td>
<td>TREC</td>
</tr>
</tbody>
</table>
“When was Alaska purchased?”

Important learned QA feature
question starts with (“when”)
AND
passage contains entity of type(DATE)
“When was Alaska purchased?”

Standard search query: (alaska, purchased)

Improved query: (alaska, purchased) & contains(DATE)
“When was Alaska purchased?”

Improved query: ( alaska, purchased ) & contains( DATE )

Results
  improved search (17% relative MRR)
  improved extraction (20% relative F1 for top candidate)
Feature driven answer extraction
Feature driven retrieval
Feature driven QA on a KB
Feature driven answer extraction
Feature driven retrieval
Feature driven QA on a KB
structured query
structured query
structured query
Question Answering on FreeBase

- Features
- Simple query
- Candidate
- Candidate
- Candidate
- A
who is the brother of Justin Bieber?
who is the brother of Justin Bieber?
who is the brother of Justin Bieber?
who is the brother of Justin Bieber?
who is the brother of Justin Bieber?
features $\rightarrow$ simple query $\rightarrow$ Justin Bieber
Freebase Topic Graph

Justin Bieber

London

awards_won

place_of_birth

Justin Bieber

dummy node

sibling

Jazmyn Bieber

gender

female

type

person

Jaxon Bieber

gender

male

type

person
Freebase Topic Graph
## Features on Graph

extract features for each node

<table>
<thead>
<tr>
<th>Justin Bieber</th>
<th>Jazmyn Bieber</th>
<th>Jaxon Bieber</th>
</tr>
</thead>
<tbody>
<tr>
<td>has:awards_won</td>
<td>has:sibling</td>
<td>has:sibling</td>
</tr>
<tr>
<td>has:place_of_birth</td>
<td>gender:female</td>
<td>gender:male</td>
</tr>
<tr>
<td>type:person</td>
<td>type:person</td>
<td>type:person</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Features on Question
Joint Features

on graph

- Justin Bieber
  - has:awards_won
  - has:place_of_birth
  - has:sibling
    - type:person
  - ...

- Jazmyn Bieber
  - has:sibling
    - gender:female
    - type:person
  - ...

- Jaxon Bieber
  - has:sibling
    - gender:male
    - type:person
  - ...

on question

for every edge e(s,t), extract:
- s, t, s|t, and s|e|t

- qword=what
- qfocus=name
- qverb=be
- qtopic=person
- qword=what|cop|qverb=be
- qword=what|nsubj|qfocus=name
- brother|nn|qtopic=person
- ...

...
Evaluation

Data: WebQuestions
Berant, et. al. (2013)
5810 questions crawled from Google Suggest
Evaluation

Data: WebQuestions

5810 questions crawled from Google Suggest

Berant et al. (2013)
Evaluation

Data: WebQuestions

5810 questions crawled from Google Suggest

Berant et al. (2013)

which states does the connecticut river flow through?
who does david james play for 2011?
what date was john adams elected president?
what kind of currency does cuba use?
who owns the cleveland browns?
...
Model

L1 regularized Logistic Regression
high-performance toolkit: Classias (Okazaki, 2009)

Start with: 7 million feature types extracted
Train for: 4 hours
Result: 30 thousand features with non-zero weight
Our Results

- Yih et al. '15
- Berant and Liang, '14
- Yao and Van Durme, '14
- Berant et al. '13

F1 Avg
Accuracy
Stanford

Yih et al. '15
Berant and Liang, '14
Yao and Van Durme, '14
Berant et al. '13

F1 Avg
Accuracy

0 10 20 30 40 50 60
Microsoft Research

Yih et al. '15
Berant and Liang, '14
Yao and Van Durme, '14
Berant et al. '13

F1 Avg
Accuracy
Feature driven answer extraction
Feature driven retrieval
Feature driven QA on a KB
Feature driven answer extraction
Feature driven retrieval
Feature driven QA on a KB
This work supported by …