Lecture 10: Question answering

LTAT.01.001 – Natural Language Processing
Kairit Sirts (kairit.sirts@ut.ee)
17.04.2019
Douglas Adams: Hitchhiker’s guide to the Galaxy

The answer to the Ultimate Question was computed by the supercomputer Deep Thought

• Question: William Wilkinson's 'An account of the principalities of Wallachia and Moldavia' inspired this author's most famous novel

• Answer: who is Bram Stoker

• https://www.youtube.com/watch?v=P18EdAKuC1U
The University of Tartu is a classical university in the city of Tartu, Estonia. It is the national university of Estonia. The University of Tartu is the only classical university in the country and also the biggest and most prestigious university in Estonia. 

**University of Tartu**

University

The University of Tartu is a classical university in the city of Tartu, Estonia. It is the national university of Estonia. The University of Tartu is the only classical university in the country and also the biggest and most prestigious university in Estonia. Wikipedia

**Address:** Ülikooli 18, 50090 Tartu

**Founder:** Gustavus Adolphus of Sweden

**Undergraduate tuition and fees:** 7,070 EUR (2016)

**Total enrollment:** 14,470 (2014)

**Colors:** White, Blue

Suggest an edit
What kind of questions

• Factoid questions

• Who founded Microsoft?
• What is the average age of the onset of autism?
• Where is TransferWise based?
Factoid QA methods

• IR-based question answering or text-based question answering
  • Relies on huge amounts of texts
  • Given a question, IR techniques extract passages directly from these documents

• Knowledge-based question answering
  • Build a semantic representation of the query
  • Use this representation to query databases of facts

• Hybrid systems combining IR and knowledge-based QA
  • Like DeepQA in IBM’s Watson
Information retrieval based QA
IR-based factoid QA

1. Question processing
2. Passage retrieval and ranking
3. Answer processing
Question processing

The main goal is to extract the **query** - keywords that should be used in the IR system to search for relevant documents.

Some systems additionally extract:

- **Answer type** - the kind of entity the answer consists of:
  - Person, location, time, etc
- **Focus** – the sequence of words in the question that will be likely replaced by the answer
- **Question type** – what kind of question it is?
  - Definition, math, list
Question processing

Which US state capital has the largest population?

Query: “US state capital has the largest population”

Answer type: city

Focus: state capital
Query formulation

- The question itself: What is the national flower of Estonia
- Leave out the question word: Is the national flower of Estonia
- Use only the content of the noun phrases: National flower of Estonia
- Remove stop-words and other high-frequency words: national flower Estonia
- Query reformulation: The national flower of Estonia is
Answer type detection

• Detect the kind of entity the answer consists of: person, location, time, etc
Answer type taxonomy

A subset from Li and Roth (2005)
Answer type detection

• Detect the kind of entity the answer consists of: person, location, time, etc

• Rule-based system using regular expressions
  • who {is | was | are | were} PERSON --> BIOGRAPHY

• Supervised classifiers:
  • Trained on databases of questions annotated with answer types
  • Can be feature based
    • Word embeddings, POS, named entities
    • **Question headword** – head of the first NP after the question’s wh-word
  • Can use neural models
Question headword

Head of the first NP after the question’s wh-word
1. Segment the retrieved documents into passages
   a) Answer type classification on passages
   b) Filter out passages with mismatching type

2. Rank the passages using supervised machine learning:
   • The number of named entities of the right type
   • The number of question keywords
   • The longest exact sequence of question keywords occurring in the passage
   • The rank of the document
   • The number of overlapping ngrams between passage and query
   • The proximity of keywords to each other in passage
Answer extraction

• Extract specific answer from the passage
• Modeled as span labeling
  • Given passage, identify the span of text that constitutes the answer
Who is the rector of University of Tartu?

The 263-member electoral council of the University of Tartu on Thursday elected Toomas Asser, professor of neurosurgery, as the university's rector for the next five years. Apr 26, 2018

Toomas Asser elected rector of University of Tartu - Estonian news
How tall is Mount Everest today?

Current Height of Mount Everest. The first measurement of Mount Everest was taken in 1852, and put the peak at **29,002 feet**. The most recent measurement clocked in at **29,029 feet**. The tectonic plates underneath the mountain account for the changes in height—the peak has jumped 27 feet in the past 165 years. Jul 5, 2017

How Tall Is Mount Everest in 2017? Why it Might Be Sinking | Travel + ...
Answer extraction based on NER

• Extract potential answers using answer type info and named entity information

• Question of type HUMAN

“Who is the rector of University of Tartu?”
The 263-member electoral council of the University of Tartu on Thursday elected Toomas Asser, professor of neurosurgery, as the university’s rector for the next five years.

• Question of type DISTANCE-QUANTITY

“How tall is Mount Everest today?”
Current height of Mount Everest. The first measurement of Mount Everest was taken in 1852 and put the peak at 29,002 feet. The most recent measurement clocked in at 29,029 feet.
Rule-based answer extraction

• The answers to some questions, such as DEFINITION, do not tend to be of particular named entity type
• Hand-written or automatically learned patterns can be used to extract these answers

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;AP&gt; such as &lt;QP&gt;</td>
<td>What is autism?</td>
<td>“, developmental disorders such as autism”</td>
</tr>
<tr>
<td>&lt;QP&gt;, a &lt;AP&gt;</td>
<td>What is a caldera?</td>
<td>“the Long Valley caldera, a <strong>volcanic crater</strong> 19 miles long”</td>
</tr>
</tbody>
</table>
Features for supervise candidate answer ranking

Does the span or sentence contain an answer? (ranking)

- **Answer type match**: candidate contains a phrase with the correct answer type
- **Pattern match**: regular expression pattern matches the candidate
- **Question keywords**: # of question keywords in the candidate
- **Keyword distance**: distance in words between the candidate and query keywords
- **Novelty factor**: a word in the candidate is not in the query
- **Apposition features**: the candidate is an appositive to question terms
- **Punctuation location**: the candidate is immediately followed by a comma, period, quotation marks, semicolon or exclamation mark
- **Sequences of question terms**: the length of the longest sequence of question terms that occurs in the candidate answer
Answer extraction by ngram tiling

• Used in web search
• Uses the snippets returned by the web search engine

1. N-gram mining
   • Extract unigrams, bigrams and trigrams from the snippets

2. N-gram filtering
   • Score n-grams by how well they match the predicted answer type

3. N-gram tiling
   • Concatenate overlapping n-gram fragments into longer answers
Neural answer extraction

• Based on the intuition that the question and the answer are in some way semantically similar
• Compute embedding representation for the query
• Compute embedding for each token in the passage
• Select passage spans whose embeddings are closest to the question embedding
Reading comprehension task

• The person/child/NLP system must read a passage and then answer questions about these passage

• SQuAD – Stanford Question Answering Dataset
  • >150000 questions
  • consists of passages from Wikipedia and associated questions
  • some questions are unanswerable
SQuAD example

Beyoncé Giselle Knowles-Carter (born September 4, 1981) is an American singer, songwriter, record producer and actress. Born and raised in Houston, Texas, she performed in various singing and dancing competitions as a child, and rose to fame in the late 1990s as lead singer of R&B girl-group Destiny’s Child. Managed by her father, Mathew Knowles, the group became one of the world’s best-selling girl groups of all time. Their hiatus saw the release of Beyoncé’s debut album, Dangerously in Love (2003), which established her as a solo artist worldwide, earned five Grammy Awards and featured the Billboard Hot 100 number-one singles “Crazy in Love” and “Baby Boy”.

| Q: “In what city and state did Beyoncé grow up?” | A: “Houston, Texas” |
| Q: “What areas did Beyoncé compete in when she was growing up?” | A: “singing and dancing” |
Neural answer extraction (Chen et al., 2017)
Neural answer extraction (Chen et al., 2017)

q - question
p – passage

• For each word token in passage p, compute the probability that this token
  • starts the answer span
  • ends the answer span
Neural answer extraction (Chen et al., 2017)

\[ q = \sum_{j} b_j q_j \]

\[ b_j = \frac{\exp(w \cdot q_j)}{\sum_{j'} \exp(w \cdot q'_{j'})} \]

question embedding

learnable parameters
Neural answer extraction (Chen et al., 2017)

- Passage embedding
  - word embedding
  - POS/NER tag embedding
  - Exact match feature – did the passage word occur in the question
  - Aligned question embedding
Neural answer extraction (Chen et al., 2017)

- Aligned question embedding

\[ \sum_j a_{i,j} \mathbf{E}(q_j) \]

\[
a_{i,j} = \frac{\exp(\alpha(\mathbf{E}(p_i)) \cdot \alpha(\mathbf{E}(q_j)))}{\sum_{j'} \exp(\alpha(\mathbf{E}(p_i)) \cdot \alpha(\mathbf{E}(q'_{j'})))}
\]
Neural answer extraction (Chen et al., 2017)

For each word token in passage p, compute the probability that this token
  • starts the answer span
  • ends the answer span

\[
\begin{align*}
  p_{\text{start}}(i) & \propto \exp(p_i W_s q) \\
  p_{\text{end}}(i) & \propto \exp(p_i W_e q)
\end{align*}
\]
Knowledge-based question answering
Knowledge-based QA

• Use a structured database to find the answer to a question
• The question in natural language must be mapped to a query to search from that database
• The query is typically some kind of logical form
• **Semantic parsers** are used to map questions to logical forms

<table>
<thead>
<tr>
<th>Question</th>
<th>Logical form</th>
</tr>
</thead>
<tbody>
<tr>
<td>When was Ada Lovelace born?</td>
<td>birth-year (Ada Lovelace, ?x)</td>
</tr>
<tr>
<td>What states border Texas?</td>
<td>( \lambda x.\text{state}(x) \land \text{borders}(x,\text{texas}) )</td>
</tr>
<tr>
<td>What is the largest state</td>
<td>( \text{argmax}(\lambda x.\text{state}(x), \lambda x.\text{size}(x)) )</td>
</tr>
<tr>
<td>How many people survived the sinking of the Titanic</td>
<td>( \text{count} (!\text{fb:event.disaster.survivors fb:en.sinking.of.the.titanic}) )</td>
</tr>
</tbody>
</table>
RDF database and queries

• Databases like Freebase and DBpedia contain large number of factual triples derived from Wikipedia infoboxes

```
subject       predicate       object
Ada Lovelace  birth-year     1815
```

“When was Ada Lovelace born?” → birth-year (Ada Lovelace, ?x)
“What is the capital of England?” → capital-city(?x, England)
Methods for mapping questions to queries

• Rule-based methods
• Supervised methods
• Semi-supervised methods
Rule-based mapping

• Useful for simple and frequent relations, such as birth-year

when * PERSON born?

birth-year(PERSON, ?x)
Supervised mapping

- Trained on a set of question-logical form pairs
  - Questions are first parsed
  - Parse trees are aligned to logical forms

- Might use a small set of seed rules
- Use these rules to bootstrap more specific alignment rules
- Might assume that all entities are known to the system
Supervised mapping - bootstrapping

- Might use a small set of simple rules such as:

```
Who V ENTITY → relation( ?x, entity)
```

```
When V ENTITY → relation( ?x, entity)
```
Learning specific rules

• Given training examples such as:

  “When was Ada Lovelace born?” \(\rightarrow\) birth-year (Ada Lovelace, ?x)

• Parse the questions and form the mappings:

  ![Diagram](image)

  When was Ada Lovelace born \(\rightarrow\) birth-year(Ada Lovelace, ?x)

• From many examples like this, infer the specific rule:

  ![Diagram](image)

  When was \(\cdot\) born \(\rightarrow\) birth-year( , ?x)
Semi-supervised mapping

• It is difficult to create supervised training sets that cover enough variations of how questions can be expressed
  • One relation can be expressed using many different phrases

• This problem can be alleviated by using unlabeled web sources

• Open Information Extraction:
  • Extract (subject, relation, object) triples of strings from web

• Align extracted triples with a canonical knowledge source, such as Wikipedia
Alignment with Freebase

Berant et al., 2013. Semantic Parsing on Freebase from Question-Answer Pairs

• A set of phrases that align with Freebase relation `country.capital`

<table>
<thead>
<tr>
<th>capital of</th>
<th>capital city of</th>
<th>become capital of</th>
</tr>
</thead>
<tbody>
<tr>
<td>capitol of</td>
<td>national capital of</td>
<td>official capital of</td>
</tr>
<tr>
<td>political capital of</td>
<td>administrative capital of</td>
<td>beautiful capital of</td>
</tr>
<tr>
<td>capital city of</td>
<td>remain capital of</td>
<td>make capital of</td>
</tr>
<tr>
<td>political center of</td>
<td>bustling capital of</td>
<td>capital city in</td>
</tr>
<tr>
<td>cosmopolitan capital of</td>
<td>move its capital to</td>
<td>modern capital of</td>
</tr>
<tr>
<td>federal capital of</td>
<td>beautiful capital city of</td>
<td>administrative capital city of</td>
</tr>
</tbody>
</table>
Alignment with paraphrase databases

Fader et al., 2013. Paraphrase-Driven Learning for Open Question Answering

• PARALEX corpus (collected from wikianswers.com)

Q: What are the green blobs in plant cells?

Lemmatized synonyms from PARALEX:
what be the green blob in plant cell?
what be green part in plant cell?
what be the green part of a plant cell?
what be the green substance in plant cell?
what be the part of plant cell that give it green color?
what cell part do plant have that enable the plant to be give a green color?
what part of the plant cell turn it green?
part of the plant cell where the cell get it green color?
the green part in a plant be call?
the part of the plant cell that make the plant green be call?
Hybrid systems
DeepQA - the QA component of IBM Watson
DeepQA

The main idea is to propose a large number of answer candidates extracted from different sources and rank them using various pieces of evidence.

Main stages:
1. Question processing
2. Candidate answer generation
3. Candidate answer scoring
4. Confidence merging and ranking
(1) Question processing

- Run parsing, NER and relation extraction on the question
- Extract **focus** and **answer type**
- Perform **question classification**
Poets and Poetry: **He** was a bank clerk in the Yukon before he published “Songs of a Sourdough” in 1907.

- **NER**
  - Yukon --> GEOPOLITICAL ENTITY
  - “Songs of a Sourdough” --> COMPOSITION

- **Coreference:**
  - He --> clerk

- **Relation extraction**
  - authorof(focus, “Songs of a Sourdough”)
  - publish(e1, he, “Songs of a Sourdough”)
  - in(e2, e1, 1907)
  - temporallink(publish(...), 1907)
• Extract **focus** with hand-written rules: **He**
  • Focus is the part of the question that co-refers with the answer

• Extract **lexical answer type** with hand-written rules: **He, poet, clerk**
  • It says something about the semantic type of the answer
  • The number of different answer types is ca 5000
  • The default rule is to choose the syntactic headword of the focus
  • Extract additional answer types using for instance co-referents of the focus or even the question category

• **Question classification** with regular expressions:
  • Definition question
  • Multiple choice
  • Puzzle
  • Fill-in-the-blank
(2) Candidate answer generation

- Generate many candidate answers
- Knowledge bases (DBPedia, IMDB, triples from OIE)

\[
\text{authorof(}\text{focus, }"\text{Songs of a sourdough}"\text{)}
\]

\[
\text{authorof(}\text{?x, }"\text{Songs of a sourdough}"\text{)}
\]

- IR from external documents
  - Retrieve short passages and extract answer candidates (anchor texts, noun phrases)
  - Query Wikipedia pages and return the titles
(3) Candidate answer scoring

- Uses many sources of evidence
  - Lexical answer type

**Answer candidate**
- Difficulty swallowing

**Lexical answer type**
- Manifestation

**DBPedia**
- Dysphagia

**WordNet**
- Symptom

Condition

Hyponymy (is-a)?
Instance-of?
Synonymy?
(3) Candidate answer scoring

- Uses many sources of evidence
  - Lexical answer type
  - Time and space relations
  - Retrieve texts with supporting evidence
(4) Answer merging and scoring

- Merge equivalent candidate answers
  - Automatically generated name dictionaries
  - Morphological parsing for common nouns
- Merge the scoring evidence of each merged answer
- Score the answers with a regularized logistic regression classifier
Evaluating factoid answers

- Mean reciprocal rank – MRR
- Assumes that the correct answers to the questions are known
- Assumes that the systems returns a ranked shortlist of answers
- Each question is scored by its reciprocal of the rank of the first correct answer
- If the ranked list does not contain any correct answer then the score for that question is 0
- The score of a system is the mean scores of all test questions
MRR - example

<table>
<thead>
<tr>
<th>Question</th>
<th>Correct answer</th>
<th>Rank of the correct answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many people survived the sinking of the Titanic?</td>
<td>706</td>
<td>2</td>
</tr>
<tr>
<td>What is the average temperature in Sydney in August?</td>
<td>14°</td>
<td>Not among the top 5</td>
</tr>
<tr>
<td>When did Mount Fuji last erupt?</td>
<td>16.12.1707</td>
<td>1</td>
</tr>
</tbody>
</table>

MRR =
MRR - example

<table>
<thead>
<tr>
<th>Question</th>
<th>Correct answer</th>
<th>Rank of the correct answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>How many people survived the sinking of the Titanic?</td>
<td>706</td>
<td>2</td>
</tr>
<tr>
<td>What is the average temperature in Sydney in August?</td>
<td>14°</td>
<td>Not among the top 5</td>
</tr>
<tr>
<td>When did Mount Fuji last erupt?</td>
<td>16.12.1707</td>
<td>1</td>
</tr>
</tbody>
</table>

MRR = (1/2 + 0 + 1/1) / 3 = 0.5