Question-Answering: Shallow & Deep Techniques for NLP

Deep Processing Techniques for NLP
Ling 571
March 12, 2014

(Examples from Dan Jurafsky)
Roadmap

- Question-Answering:
  - Definitions & Motivation

- Basic pipeline:
  - Question processing
  - Retrieval
  - Answering processing

- Shallow processing: Aranea (Lin, Brill)

- Deep processing: LCC (Moldovan, Harabagiu, et al)

- Wrap-up
Why QA?

- Grew out of information retrieval community
- Document retrieval is great, but...
  - Sometimes you don’t just want a ranked list of documents
  - Want an answer to a question!
    - Short answer, possibly with supporting context
- People ask questions on the web
  - Web logs:
    - Which English translation of the bible is used in official Catholic liturgies?
    - Who invented surf music?
    - What are the seven wonders of the world?
  - Account for 12-15% of web log queries
Search Engines and Questions

- What do search engines do with questions?
  - Increasingly try to answer questions
    - Especially for wikipedia infobox types of info
  - Backs off to keyword search

- How well does this work?
  - Who invented surf music?
    - Rank #2 snippet:
      - Dick Dale *invented surf music*
    - Pretty good, but...
Search Engines & QA

- **Who was the prime minister of Australia during the Great Depression?**
  - Rank 1 snippet:
    - The conservative *Prime Minister of Australia*, Stanley Bruce
  - Wrong!
    - Voted out just before the Depression

- **What is the total population of the ten largest capitals in the US?**
  - Rank 1 snippet:
    - The table below lists the **largest 50 cities in the United States** .....  
  - The answer is in the document – with a calculator..
Search Engines and QA

- Search for exact question string
  - “Do I need a visa to go to Japan?”
    - Result: Exact match on Yahoo! Answers
    - Find ‘Best Answer’ and return following chunk

- Works great if the question matches exactly
  - Many websites are building archives
  - What if it doesn’t match?
    - ‘Question mining’ tries to learn paraphrases of questions to get answer
Perspectives on QA

- TREC QA track (~2000-)
  - Initially pure factoid questions, with fixed length answers
    - Based on large collection of fixed documents (news)
    - Increasing complexity: definitions, biographical info, etc
      - Single response
  - Reading comprehension (Hirschman et al, 2000-)
    - Think SAT/GRE
      - Short text or article (usually middle school level)
      - Answer questions based on text
    - Also, ‘machine reading’
  - And, of course, Jeopardy! and Watson
<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where is the Louvre Museum located?</td>
<td>in Paris, France</td>
</tr>
<tr>
<td>What’s the abbreviation for limited partnership?</td>
<td>L.P.</td>
</tr>
<tr>
<td>What are the names of Odin’s ravens?</td>
<td>Huginn and Muninn</td>
</tr>
<tr>
<td>What currency is used in China?</td>
<td>the yuan</td>
</tr>
<tr>
<td>What kind of nuts are used in marzipan?</td>
<td>almonds</td>
</tr>
<tr>
<td>What instrument does Max Roach play?</td>
<td>drums</td>
</tr>
<tr>
<td>What’s the official language of Algeria?</td>
<td>Arabic</td>
</tr>
<tr>
<td>What is the telephone number for the University of Colorado, Boulder?</td>
<td>(303)492-1411</td>
</tr>
<tr>
<td>How many pounds are there in a stone?</td>
<td>14</td>
</tr>
</tbody>
</table>
Basic Strategy

- Given an indexed document collection, and
- A question:
- Execute the following steps:
  - Query formulation
  - Question classification
  - Passage retrieval
  - Answer processing
  - Evaluation
Query Processing

• Query reformulation
  • Convert question to suitable form for IR
    • E.g. ‘stop structure’ removal:
      • Delete function words, q-words, even low content verbs

• Question classification
  • Answer type recognition
    • Who \(\rightarrow\) Person; What Canadian city \(\rightarrow\) City
    • What is surf music \(\rightarrow\) Definition
  • Train classifiers to recognize expected answer type
    • Using POS, NE, words, synsets, hyper/hypo-nyms
<table>
<thead>
<tr>
<th>Tag</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABBREVIATION</td>
<td></td>
</tr>
<tr>
<td>abb</td>
<td>What’s the abbreviation for limited partnership?</td>
</tr>
<tr>
<td>exp</td>
<td>What does the “c” stand for in the equation E=mc2?</td>
</tr>
<tr>
<td>DESCRIPTION</td>
<td></td>
</tr>
<tr>
<td>definition</td>
<td>What are tannins?</td>
</tr>
<tr>
<td>description</td>
<td>What are the words to the Canadian National anthem?</td>
</tr>
<tr>
<td>manner</td>
<td>How can you get rust stains out of clothing?</td>
</tr>
<tr>
<td>reason</td>
<td>What caused the Titanic to sink?</td>
</tr>
<tr>
<td>ENTITY</td>
<td></td>
</tr>
<tr>
<td>animal</td>
<td>What are the names of Odin’s ravens?</td>
</tr>
<tr>
<td>body</td>
<td>What part of your body contains the corpus callosum?</td>
</tr>
<tr>
<td>color</td>
<td>What colors make up a rainbow?</td>
</tr>
<tr>
<td>creative</td>
<td>In what book can I find the story of Aladdin?</td>
</tr>
<tr>
<td>currency</td>
<td>What currency is used in China?</td>
</tr>
<tr>
<td>disease/medicine</td>
<td>What does Salk vaccine prevent?</td>
</tr>
<tr>
<td>event</td>
<td>What war involved the battle of Chapultepec?</td>
</tr>
<tr>
<td>food</td>
<td>What kind of nuts are used in marzipan?</td>
</tr>
<tr>
<td>instrument</td>
<td>What instrument does Max Roach play?</td>
</tr>
<tr>
<td>lang</td>
<td>What’s the official language of Algeria?</td>
</tr>
<tr>
<td>letter</td>
<td>What letter appears on the cold-water tap in Spain?</td>
</tr>
<tr>
<td>other</td>
<td>What is the name of King Arthur’s sword?</td>
</tr>
<tr>
<td>plant</td>
<td>What are some fragrant white climbing roses?</td>
</tr>
<tr>
<td>product</td>
<td>What is the fastest computer?</td>
</tr>
<tr>
<td>religion</td>
<td>What religion has the most members?</td>
</tr>
<tr>
<td>sport</td>
<td>What was the name of the ball game played by the Mayans?</td>
</tr>
<tr>
<td>substance</td>
<td>What fuel do airplanes use?</td>
</tr>
<tr>
<td>symbol</td>
<td>What is the chemical symbol for nitrogen?</td>
</tr>
<tr>
<td>technique</td>
<td>What is the best way to remove wallpaper?</td>
</tr>
<tr>
<td>term</td>
<td>How do you say “Grandma” in Irish?</td>
</tr>
<tr>
<td>vehicle</td>
<td>What was the name of Captain Bligh’s ship?</td>
</tr>
<tr>
<td>word</td>
<td>What’s the singular of dice?</td>
</tr>
<tr>
<td>Category</td>
<td>Question</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>HUMAN</td>
<td>Who was Confucius?</td>
</tr>
<tr>
<td></td>
<td>What are the major companies that are part of Dow Jones?</td>
</tr>
<tr>
<td></td>
<td>Who was the first Russian astronaut to do a spacewalk?</td>
</tr>
<tr>
<td></td>
<td>What was Queen Victoria's title regarding India?</td>
</tr>
<tr>
<td>LOCATION</td>
<td>What's the oldest capital city in the Americas?</td>
</tr>
<tr>
<td></td>
<td>What country borders the most others?</td>
</tr>
<tr>
<td></td>
<td>What is the highest peak in Africa?</td>
</tr>
<tr>
<td></td>
<td>What river runs through Liverpool?</td>
</tr>
<tr>
<td></td>
<td>What states do not have state income tax?</td>
</tr>
<tr>
<td>NUMERIC</td>
<td>What is the telephone number for the University of Colorado?</td>
</tr>
<tr>
<td></td>
<td>About how many soldiers died in World War II?</td>
</tr>
<tr>
<td></td>
<td>What is the date of Boxing Day?</td>
</tr>
<tr>
<td></td>
<td>How long was Mao's 1930s Long March?</td>
</tr>
<tr>
<td></td>
<td>How much did a McDonald's hamburger cost in 1963?</td>
</tr>
<tr>
<td></td>
<td>Where does Shanghai rank among world cities in population?</td>
</tr>
<tr>
<td></td>
<td>What is the population of Mexico?</td>
</tr>
<tr>
<td></td>
<td>What was the average life expectancy during the Stone Age?</td>
</tr>
<tr>
<td></td>
<td>What fraction of a beaver's life is spent swimming?</td>
</tr>
<tr>
<td></td>
<td>What is the speed of the Mississippi River?</td>
</tr>
<tr>
<td></td>
<td>How fast must a spacecraft travel to escape Earth's gravity?</td>
</tr>
<tr>
<td></td>
<td>What is the size of Argentina?</td>
</tr>
<tr>
<td></td>
<td>How many pounds are there in a stone?</td>
</tr>
</tbody>
</table>
Passage Retrieval

- Why not just perform general information retrieval?
  - Documents too big, non-specific for answers

- Identify shorter, focused spans (e.g., sentences)
  - Filter for correct type: answer type classification
  - Rank passages based on a trained classifier

- Or, for web search, use result snippets
Answer Processing

- Find the specific answer in the passage
- Pattern extraction-based:
  - Include answer types, regular expressions
  - Can use syntactic/dependency/semantic patterns
  - Leverage large knowledge bases

<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 \textcolor{red}{volcanic crater} 19 miles long”</td>
</tr>
</tbody>
</table>
Evaluation

- Classical:
  - Return ranked list of answer candidates
  - Idea: Correct answer higher in list => higher score

- Measure: Mean Reciprocal Rank (MRR)
  - For each question,
    - Get reciprocal of rank of first correct answer
      - E.g. correct answer is 4 => $\frac{1}{4}$
      - None correct => 0
    - Average over all questions

\[
MRR = \frac{\sum_{i=1}^{N} \frac{1}{\text{rank}_i}}{N}
\]
AskMSR/Aranea (Lin, Brill)

- Shallow Processing for QA

1. Question
   Where is the Louvre Museum located?

2. Rewrite Query
   “+the Louvre Museum +is located”

3. <Search Engine>

4. Collect Summaries, Mine N-grams

5. N-Best Answers

6. Tile N-Grams

7. Filter N-Grams
Intuition

- Redundancy is useful!
  - If similar strings appear in many candidate answers, likely to be solution
    - Even if can’t find obvious answer strings

- Q: How many times did Bjorn Borg win Wimbledon?
  - Bjorn Borg blah blah blah Wimbledon blah 5 blah
  - Wimbledon blah blah blah Bjorn Borg blah 37 blah.
  - blah Bjorn Borg blah blah 5 blah blah Wimbledon
  - 5 blah blah Wimbledon blah blah Bjorn Borg.
  - Probably 5
Query Reformulation

- Identify question type:
  - E.g. Who, When, Where,...

- Create question-type specific rewrite rules:
  - Hypothesis: Wording of question similar to answer
    - For ‘where’ queries, move ‘is’ to all possible positions
      - Where is the Louvre Museum located? =>
        - Is the Louvre Museum located
        - The is Louvre Museum located
        - The Louvre Museum is located, .etc.

- Create type-specific answer type (Person, Date, Loc)
Retrieval, N-gram Mining & Filtering

- Run reformulated queries through search engine
  - Collect (lots of) result snippets

- Collect n-grams from snippets

- Weight each n-gram summing over occurrences

- Concatenate n-grams into longer answers
  - E.g. Dickens, Charles Dickens, Mr. Charles
  - Mr. Charles Dickens
Example Redux

Question
Where is the Louvre Museum located?

in Paris France 59%
museums 12%
hostels 10%

N-Best Answers

Rewrite Query
“+the Louvre Museum +is located”
“+the Louvre Museum +is +in”
“+the Louvre Museum +is near”
“+the Louvre Museum +is”
Louvre AND Museum AND near

<Search Engine>

Collect Summaries, Mine N-grams

Filter N-grams

Tile N-Grams
Deep Processing Technique for QA

- LCC, PowerAnswer, Qanda (Moldovan, Harabagiu, et al)
Deep Processing: Query/Answer Formulation

- Preliminary shallow processing:
  - Tokenization, POS tagging, NE recognition, Preprocess

- Parsing creates syntactic representation:
  - Focused on nouns, verbs, and particles
    - Attachment

- Coreference resolution links entity references

- Translate to full logical form
  - As close as possible to syntax
Syntax to Logical Form

Diagram showing the transformation from syntax to logical form.
Deep Processing:
Answer Selection

- Lexical chains:
  - Bridge gap in lexical choice b/t Q and A
    - Improve retrieval and answer selection
  - Create connections via WordNet synsets
    - Q: When was the internal combustion engine invented?
    - A: The first internal-combustion engine was built in 1867.
    - invent → create_mentally → create → build

- Perform abductive reasoning
  - Tries to justify answer given question
  - Yields 30% improvement in accuracy!
Question Answering Example

- How hot does the inside of an active volcano get?
  - get(TEMPERATURE, inside(volcano(active)))
Question Answering Example

- How hot does the inside of an active volcano get?
  - get(TEMPERATURE, inside(volcano(active)))

- “lava fragments belched out of the mountain were as hot as 300 degrees Fahrenheit”
  - fragments(lava, TEMPERATURE(degrees(300)), belched(out, mountain))

- volcano ISA mountain
- lava ISPARTOF volcano ■ lava inside volcano
- fragments of lava HAVEPROPERTIESOF lava
Question Answering Example

• How hot does the inside of an active volcano get?
  • get(TEMPERATURE, inside(volcano(active)))

• “lava fragments belched out of the mountain were as hot as 300 degrees Fahrenheit”
  • fragments(lava, TEMPERATURE(degrees(300)), belched(out, mountain))
  • volcano ISA mountain
  • lava ISPARTOF volcano ■ lava inside volcano
  • fragments of lava HAVEPROPERTIESOF lava

• The needed semantic information is in WordNet definitions, and was successfully translated into a form that was used for rough ‘proofs’
A Victory for Deep Processing

<table>
<thead>
<tr>
<th>Run Tag</th>
<th>Score</th>
<th>#</th>
<th>%</th>
<th>Inexact</th>
<th>Prec</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>LCCmain2002</td>
<td>0.856</td>
<td>415</td>
<td>83.0</td>
<td>8</td>
<td>0.578</td>
<td>0.804</td>
</tr>
<tr>
<td>exactanswer</td>
<td>0.691</td>
<td>271</td>
<td>54.2</td>
<td>12</td>
<td>0.222</td>
<td>0.848</td>
</tr>
<tr>
<td>pris2002</td>
<td>0.610</td>
<td>290</td>
<td>58.0</td>
<td>17</td>
<td>0.241</td>
<td>0.891</td>
</tr>
<tr>
<td>IRST02D1</td>
<td>0.589</td>
<td>192</td>
<td>38.4</td>
<td>17</td>
<td>0.167</td>
<td>0.217</td>
</tr>
<tr>
<td>IBMPQSQACYC</td>
<td>0.588</td>
<td>179</td>
<td>35.8</td>
<td>9</td>
<td>0.196</td>
<td>0.630</td>
</tr>
<tr>
<td>uwmtB3</td>
<td>0.512</td>
<td>184</td>
<td>36.8</td>
<td>20</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>BBN2002C</td>
<td>0.499</td>
<td>142</td>
<td>28.4</td>
<td>18</td>
<td>0.182</td>
<td>0.087</td>
</tr>
<tr>
<td>isi02</td>
<td>0.498</td>
<td>149</td>
<td>29.8</td>
<td>15</td>
<td>0.385</td>
<td>0.109</td>
</tr>
<tr>
<td>limsiQalir2</td>
<td>0.497</td>
<td>133</td>
<td>26.6</td>
<td>11</td>
<td>0.188</td>
<td>0.196</td>
</tr>
<tr>
<td>ali2002b</td>
<td>0.496</td>
<td>181</td>
<td>36.2</td>
<td>15</td>
<td>0.156</td>
<td>0.848</td>
</tr>
<tr>
<td>ibmsqa02c</td>
<td>0.455</td>
<td>145</td>
<td>29.0</td>
<td>44</td>
<td>0.224</td>
<td>0.239</td>
</tr>
<tr>
<td>FDUT11QA1</td>
<td>0.434</td>
<td>124</td>
<td>24.8</td>
<td>6</td>
<td>0.139</td>
<td>0.957</td>
</tr>
<tr>
<td>aranea02a</td>
<td>0.433</td>
<td>152</td>
<td>30.4</td>
<td>36</td>
<td>0.235</td>
<td>0.174</td>
</tr>
<tr>
<td>nuslamp2002</td>
<td>0.396</td>
<td>105</td>
<td>21.0</td>
<td>17</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Aranea: 0.30 on TREC data; 0.42 on TREC queries w/full web
Conclusions

- Deep processing for QA
  - Exploits parsing, semantics, anaphora, reasoning
  - Computationally expensive
    - But tractable because applied only to
    - Questions and Passages

- Trends:
  - Systems continue to make greater use of
    - Web resources: Wikipedia, answer repositories
    - Machine learning!!!
Summary

- Deep processing techniques for NLP
  - Parsing, semantic analysis, logical forms, reference, etc
  - Create richer computational models of natural language
    - Closer to language understanding

- Shallow processing techniques have dominated many areas
  - IR, QA, MT, WSD, etc
    - More computationally tractable, fewer required resources

- Deep processing techniques experiencing resurgence
  - Some big wins – e.g. QA
  - Improved resources: treebanks (syn/disc, Framenet, Propbank)
  - Improved learning algorithms: structured learners, ...
  - Increased computation: cloud resources, Grid, etc
Notes

• Last assignment posted – Due March 18

• Course evaluation web page posted:
  • Under ‘Course Resources’ on syllabus page
  • Please respond!

• THANK YOU!