Logistics

• Project progress discussion meeting with instructor
  • Meetings on March 29 morning will take place in Rm 110, WVH (our classroom).
  • Meetings on March 29 afternoon will take place in Rm 911, 177 Huntington Ave.

Project presentation

• Each team will present for 10 minutes, with 2 minutes for QA.

• After all representations, we will vote for favorite project. Each team has two votes.

• The team that gets the most votes wins. Each team member will get 1% bonus towards the final grade.

• One winning team is selected on each of April 9, 12, and 16.

Presentation and final report

• Problem Description (10%)
  What is the task?
  System input and output
  Examples will be helpful

• Reference/Related work (20%)
  Put your work in context: what has been done before? You need to have reference!
  What’s new in your work?

• Methodology: What you have done (30%)
  Preprocessing of the data
  What are your data? Features used? What are effective, and what are not?
  What methods do you experiment with? And why do you think they’re reasonable and suitable for the task?

• Experiments (40%)
  Datasets size, train/test/development
  Evaluation metrics: what are used and are they proper to calibrate system performance?
  Baselines: what are they?
  Results, tables, figures, etc.

Question Answering
Questions in Search

Google

Question Answering (Some Background)
One of the oldest NLP tasks (punched card systems in 1961)

Question: What do worms eat?

Potential Answers:
- Worms eat grass
- Horses eat worms eat grass
- Grass is eaten by worms

WILLIAM WILKINSON’S “AN ACCOUNT OF THE PRINCIPALITIES OF WALLACHIA AND MOLDOVIA” INSPIRED THIS AUTHOR’S MOST FAMOUS NOVEL Bram Stoker

Question Answering: IBM’s Watson

• Won Jeopardy on February 16, 2011!
Types of Questions in Modern Systems

- Factoid questions
  - Who wrote “The Universal Declaration of Human Rights”?
  - How many calories are there in two slices of apple pie?
  - What is the average age of the onset of autism?
  - Where is Apple Computer based?
- Complex (narrative) questions:
  - In children with an acute febrile illness, what is the efficacy of acetaminophen in reducing fever?
  - What do scholars think about Jefferson's position on dealing with pirates?

Commercial systems: mainly factoid questions

<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>
</tbody>
</table>

Paradigms for Factoid QA

- Information Retrieval (IR)-based approaches
  - IBM Watson (some parts); Google
- Knowledge-based and Hybrid approaches
  - IBM Watson; Apple Siri; Wolfram Alpha
- Built upon the above two:
  - Data-driven, neural network-based approaches (more recent)

Information Retrieval (IR)-based QA

- Factoid QA pipeline
  - Factoid QA evaluation
  - Common Knowledge used in QA
  - Recent QA tasks
Question Processing: Things to extract from the question

- **Answer Type Detection**
  - Decide the named entity type (person, place) of the answer
- **Query Formulation**
  - Choose query keywords for the IR system
- **Question Type classification**
  - Is this a definition question, a math question, a list question?
- **Focus Detection**
  - Find the question words that are replaced by the answer
- **Relation Extraction (if there are more than one entities)**
  - Find relations between entities in the question

**Example:**

Jeopardy: They're the two states you could be reentering if you're crossing Florida's northern border.

- **Answer Type:** US state
- **Query Formulation:** two states, border, Florida, north
- **Focus:** the two states
- **Relations:** borders(Florida, ?x, north)

IR-based Factoid QA

- **QUESTION PROCESSING**
  - Detect question type, answer type, focus, relations
  - Formulate queries to send to a search engine
    - "Who is the president of US?" - person
- **PASSAGE RETRIEVAL**
  - Retrieve ranked documents
  - Break into suitable passages and rerank
- **ANSWER PROCESSING**
  - Extract candidate answers
  - Rank candidates
  - Using evidence from the text and external sources

**Example:**

Who founded Virgin Airlines?

- **Question:** Who founded Virgin Airlines?

Answer Type Detection: Named Entities

- **Who founded Virgin Airlines?**
Answer Type Detection: Named Entities

• Who founded Virgin Airlines?
  • PERSON
• What Canadian city has the largest population?
  • CITY

Answer Type Taxonomy

6 coarse classes
• ABBREVIATION, ENTITY, DESCRIPTION, HUMAN, LOCATION, NUMERIC

50 finer classes
• LOCATION: city, country, mountain...
• HUMAN: group, individual, title, description...
• ENTITY: animal, body, color, currency...

Part of Li & Roth’s Answer Type Taxonomy

Answer Types

<table>
<thead>
<tr>
<th>ENTITY</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
<td>What are the names of Sina’s cows?</td>
</tr>
<tr>
<td>body</td>
<td>What part of your body contains the unique collision?</td>
</tr>
<tr>
<td>color</td>
<td>What colors make up a rainbow?</td>
</tr>
<tr>
<td>causes</td>
<td>In what book can I find the story of Atlantis?</td>
</tr>
<tr>
<td>dates</td>
<td>What was the date of the invention of computer?</td>
</tr>
<tr>
<td>event</td>
<td>What was the result of the battle of Chouzhou?</td>
</tr>
<tr>
<td>food</td>
<td>What kind of meals are served on board?</td>
</tr>
<tr>
<td>instrument</td>
<td>What instrument does Moka Business use?</td>
</tr>
<tr>
<td>lang</td>
<td>What is the official language of Alberta?</td>
</tr>
<tr>
<td>story</td>
<td>What time appears on the clock when it is 9 o'clock?</td>
</tr>
<tr>
<td>sport</td>
<td>What sport did the Greeks invent?</td>
</tr>
<tr>
<td>plant</td>
<td>What are some important plants containing roses?</td>
</tr>
<tr>
<td>product</td>
<td>What is the famous computer?</td>
</tr>
<tr>
<td>religion</td>
<td>What religion has the most members?</td>
</tr>
<tr>
<td>poet</td>
<td>What was the name of the bull god played by the Myrians?</td>
</tr>
<tr>
<td>religion</td>
<td>What is the religion of the goddess?</td>
</tr>
<tr>
<td>symbol</td>
<td>What is the symbol for the element?</td>
</tr>
<tr>
<td>chemical</td>
<td>What is the chemical symbol for the substance?</td>
</tr>
<tr>
<td>unit</td>
<td>How do you say “Gandalf” in Elvish?</td>
</tr>
<tr>
<td>weight</td>
<td>What was the name of Captain Noodle’s ship?</td>
</tr>
</tbody>
</table>

More Answer Types

<table>
<thead>
<tr>
<th>NUMERIC</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>code</td>
<td>What is the telephone number for the University of Colorado?</td>
</tr>
<tr>
<td>cost</td>
<td>About how much money did the World War II cost?</td>
</tr>
<tr>
<td>date</td>
<td>What is the date on the clock?</td>
</tr>
<tr>
<td>food</td>
<td>What is the food that the French eat?</td>
</tr>
<tr>
<td>distance</td>
<td>What is the distance that the trip is?</td>
</tr>
<tr>
<td>weight</td>
<td>What was the average life expectancy during the Stone Age?</td>
</tr>
<tr>
<td>percent</td>
<td>What percentage of Americans are of foreign origin?</td>
</tr>
<tr>
<td>size</td>
<td>What is the speed of the Mississippi River?</td>
</tr>
<tr>
<td>speed</td>
<td>How many pounds of sand is a ton?</td>
</tr>
</tbody>
</table>
Answer types in Jeopardy

- 2500 answer types in 20,000 Jeopardy question sample
- The most frequent 200 answer types cover ~ 50% of data
- The 40 most frequent Jeopardy answer types
  - country, city, man, film, state, author, group, here, company, president, capital, star, novel, character, woman, river, island, king, song, part, series, sport, singer, actor, play, team, show, actress, animal, presidential, composer, musical, nation, book, title, leader, game

Answer Type Detection

- Hand-written rules
- Machine Learning

Answer Type Detection

- Regular expression-based rules can get some cases:
  - Who (is|was|are|were) PERSON
  - PERSON (YEAR – YEAR)
- Other rules use the question headword:
  - (the headword of the first noun phrase after the wh-word)
  - Which city in China has the largest number of foreign financial companies?
  - What is the state flower of California?

Answer Type Detection

- Most often, we treat the problem as machine learning classification
  - Define a taxonomy of question types
  - Annotate training data for each question type
  - Train classifiers for each question class using a rich set of features.
    - features include those hand-written rules!

Features for Answer Type Detection

- Question words and phrases
- Part-of-speech tags
- Parse features (headwords)
- Named Entities
- Semantically related words

Which city in China has the largest number of foreign financial companies?
What is the state flower of California?

Query Formulation

- QUESTION PROCESSING
  - Detect question type, answer type, focus, relations
  - “Who is the president of US?” – person
  - Formulate queries to send to a search engine
    - “president of united states”
- PASSAGE RETRIEVAL
  - Retrieve ranked documents
  - Break into suitable passages and rerank
- ANSWER PROCESSING
  - Extract candidate answers
  - Rank candidates
  - using evidence from the text and external sources
Keyword Selection Algorithm

1. Select all non-stop words in quotations
2. Select all NNP words in recognized named entities
3. Select all complex nominals with their adjectival modifiers
4. Select all other complex nominals
5. Select all nouns with their adjectival modifiers
6. Select all other nouns
7. Select all verbs
8. Select all adverbs
9. Select the question focus word (skipped in all previous steps)
10. Select all other words

Choosing keywords from the query

Who coined the term “cyberspace” in his novel “Neuromancer”?

cyberspace/1 Neuromancer/1 term/4 novel/4 coined/7

IR-based Factoid QA

• QUESTION PROCESSING
  • Select question type, answer type, relations
  • “Who is the president of US?”
  • “Formulate queries to send to a search engine
  • “President of coined text”
• PASSAGE RETRIEVAL
  • Retrieve ranked documents
  • Break into suitable passages and rerank
• ANSWER PROCESSING
  • Extract candidate answers
  • Rank candidates
  • Using evidence from the text and external sources

Passage Retrieval

• Step 1: IR engine retrieves documents using query terms
• Step 2: Segment the documents into shorter units
  • E.g. paragraphs or consecutive sentences
• Step 3: Passage ranking
  • Use answer type to help rerank passages

Features for Passage Ranking

Either in rule-based classifiers or with supervised machine learning

• Number of Named Entities of the right type in passage
• Number of query words in passage
• Number of question N-grams also in passage
• Proximity of query keywords to each other in passage
• Longest sequence of question words
• Rank of the document containing passage

Passage Retrieval as Query-focused Summarization

Which country has the largest part of the Amazon rain forest?

[The chaotic development that is gobbling up the Amazon rain forest could finally be reined in with a new plan developed by leading scientists from around the world. “That’s one of the most encouraging signs about the Amazon rain forest in recent years,” said Thomas Lovejoy, an Amazon specialist. “It comes much later than a year ago, when there was nothing to read about conservation in the Amazon.”

[Nearly half of the Amazon, the world’s largest tropical rain forest, lies in Brazil.]

Extract passages that best summarize each document w.r.t. the query]
Passage Retrieval as Query-focused Summarization

- Decide on a summary length (10% of document length).
- Use standard ad-hoc retrieval algorithm to retrieve top k documents.
- Treat each sentence/paragraph in top N documents as a document itself.
- Use standard document similarity equations to assign a similarity score to the sentence/paragraph.
- Return highest-scoring sentences/paragraphs as the summary, subject to the length constraint.

IR-based Factoid QA

- QUESTION PROCESSING
  - Detect question type, answer type, focus, relations
  - "Who is the president of US?" - person
  - Formulate queries to send to a search engine
    - "president of United States"
- PASSAGE RETRIEVAL
  - Retrieve ranked documents
  - Break into suitable passages and rerank
- ANSWER PROCESSING
  - Extract candidate answers
  - Rank candidates
  - using evidence from the text and external sources

Answer Extraction

- Run an answer-type named-entity tagger on the passages
  - Each answer type requires a named-entity tagger that detects it
  - If answer type is GTY, tagger has to tag GTY
- Return the string with the right type:
  - Who is the prime minister of India (PERSON)
  - Manmohan Singh, Prime Minister of India, had told left leaders that the deal would not be renegotiated.
  - How tall is Mt. Everest? (LENGTH)
  - The official height of Mount Everest is 29,035 feet

Adding Analysis Patterns

- "Who is Elvis?"
  - Question type: "who"
  - Named-entity tagging: "Who is <person-name> Elvis/<person-name>"
  - Analysis pattern: if question type = "who" and question contains <person-name> then
    - Desired answer probably is a description
  - Likely answer extraction patterns
    - "Elvis, the X", e.g., "Elvis, the king of rock and roll!"
    - "the X Elvis", e.g., "the legendary entertainer Elvis"
IR-based Factoid QA

- **QUESTION PROCESSING**
  - Detect question type, answer type, focus, relations
    - “Who is the president of US?” → person
  - Formulate queries to send to a search engine
    - “president of United States”

- **PASSAGE RETRIEVAL**
  - Retrieve ranked documents
  - Break into suitable passages and rerank

- **ANSWER PROCESSING**
  - Extract candidate answers
  - Rank candidates
    - using evidence from the text and external sources

Ranking Candidate Answers

- But what if there are multiple candidate answers!

**Q**: Who was Queen Victoria’s second son?

- **Answer Type**: Person

**Passage**:
The Marie biscuit is named after Marie Alexandrovna, the daughter of Czar Alexander II of Russia and wife of Alfred, the second son of Queen Victoria and Prince Albert.

Use machine learning:
Features for ranking candidate answers

- **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.

Candidate Answer scoring in IBM Watson

- Each candidate answer gets scores from >50 components
  - (from unstructured text, semi-structured text, triple stores)
  - logical form (parse) match between question and candidate
  - passage source reliability
  - geospatial location
    - "California is “southwest of Montana”"
  - temporal relationships
  - taxonomic classification

Information Retrieval (IR)-based QA

- Factoid QA pipeline
- Factoid QA evaluation
- Common Knowledge used in QA
- Recent QA tasks
**Common Evaluation Metrics**

1. **Accuracy** (does answer match gold-labeled answer?)
2. **Mean Reciprocal Rank**
   - For each query return a ranked list of M candidate answers.
   - Query score is $1 / \text{Rank of the first correct answer}$
   - If first answer is correct: 1
   - Else if second answer is correct: $\frac{1}{2}$
   - Else if third answer is correct: $\frac{1}{3}$, etc.
   - Score is 0 if none of the M answers are correct
   - Take the mean over all N queries

\[
MRR = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\text{rank}_i}
\]

**Information Retrieval (IR)-based QA**

- Factoid QA pipeline
- Factoid QA evaluation
- Common Knowledge used in QA
- Recent QA tasks

**Knowledge in QA**

- What are other types of knowledge useful for a QA system?
  - Relations
  - Temporal information
  - Dialogue context

**Relation Extraction**

- Answers: Databases of Relations
  - born-in("Emma Goldman", "June 27 1869")
  - author-of("Cao Xue Qin", "Dream of the Red Chamber")
  - Draw from Wikipedia infoboxes, DBpedia, FreeBase, etc.
- Questions: Extracting Relations in Questions
  - Whose granddaughter starred in E.T.?
    - (acted-in ?x "E.T."
    - (granddaughter-of ?x ?y)

**Temporal Reasoning**

- Relation databases
  - (and obituaries, biographical dictionaries, etc.)
- IBM Watson
  - “In 1594 he took a job as a tax collector in Andalusia”
  - Candidates:
    - Thoreau is a bad answer (born in 1817)
    - Cervantes is possible (was alive in 1594)

**Context and Conversation in Virtual Assistants like Siri**

- Coreference helps resolve ambiguities
  - U: “Book a table at Il Fornaio at 7:00 with my mom”
  - U: “Also send her an email reminder”
- Clarification questions:
  - U: “Chicago pizza”
  - S: “Did you mean pizza restaurants in Chicago or Chicago-style pizza?”
Limitations of Factoid Q/A

- Question must query a specific fact that is explicitly stated somewhere in the document corpus.
- Does not allow aggregating or accumulating information across multiple information sources.
- Does not require “deep compositional” semantics, nor inferential reasoning to generate answer.

Information Retrieval (IR)-based QA

- Factoid QA pipeline
- Factoid QA evaluation
- Common Knowledge used in QA
- Recent QA tasks

What are recent tasks for QA?

- Reading comprehension
- Visual Question Answering

Reading Comprehension Q/A

- Answer questions that test comprehension of a specific document.
- Use standardized tests of reading comprehension to evaluate performance (Hirschman et al. 1999; Rilo & Thelen, 2000; Ng et al. 2000; Charniak et al. 2000).

Sample Reading Comprehension Test

Large Scale Reading Comprehension Data

- DeepMind’s large-scale data for reading comprehension Q/A (Hermann et al., 2015).
- News articles used as source documents.
- Questions constructed automatically from article summary sentences.
Sample DeepMind Reading Comprehension Test

Deep LSTM Reader
- DeepMind uses LSTM recurrent neural net (RNN) to encode document and query into a vector that is then used to predict the answer.

Incorporated various forms of attention to focus the reader on answering the question while reading the document.

Visual Question Answering (VQA)
- Answer natural language questions about information in images.
- VaTech/MSR group has put together VQA dataset with ~750K questions over ~250K images (Antol et al., 2016).

VQA Examples
- Answer natural language questions about information in images.
  - VaTech/MSR group has put together VQA dataset with ~750K questions over ~250K images (Antol et al., 2016).

LSTM System for VQA
- Visualize word embedding and sentence encoding.