CS544: Question Classification

February 23, 2012

Zornitsa Kozareva
USC/ISI
Marina del Rey, CA
kozareva@isi.edu
www.isi.edu/~kozareva

What do you do when you have a question?

Q> How do you get a perfect sun tan?

Search Engine

The Perfect Suntan Guide
www.maxim.com/loose-woods/the-perfect-suntan-guide.htm
Apr 21, 2006 — Don’t get burned off on holiday — too much sun is more dangerous than crime and terrorism. Follow our handy tips for a perfect reflection.

How to Get a Good Tan Without Getting Sunburned: 10 steps —
www.ew.com ↓ 13 Skin Care ↓ Sun and Sunless Tanning
Dec 17, 2011 — How to Get a Good Tan Without Getting Sunburned — Apply moisturizer or after sun and don’t scratch the suntan or it will make it worse. A ...

How to Get the Perfect Tan | ehow.com
www.ehow.com » Beauty & Personal Care
Since we know more about the dangers of worshipping the sun, such as the risks of developing skin cancer and wrinkles, getting the perfect tan now includes ...

How to get a great suntan
www.bradypodolski.com/how-to-get-a-great-suntan.html
3 votes
Exfoliating your skin is a very good idea. When you are getting a suntan the color that you get is a result of the skin reacting to the sun. If you have dead skin on ...
What do you do when you have a question?

Q> How do you get a perfect sun tan?

A> Lie in the sun

Where do we stand today

• Modern search engines provide document retrieval facilities
• (+) Given a document collection and a list of key words, they return a ranked list of relevant documents (snippets)
• (-) However
  – search engines do not speak our language
  – a document is not an answer to a question
  – relevant documents might not have the answer
  – user spends too much time to find the answer
Question Answering (QA)

- Ask the computer a question in natural language and quickly receive the exact answer (and sufficient context to validate the answer)

Ex. **Question:** Where was Einstein born?

**Answer:** Ulm

**Answer found:** One day, from among his city views of Ulm, Otto chose a water color to send to Albert Einstein as a remembrance of Einstein’s birthplace.

Simplified QA Architecture

[Diagram showing the flow of information from the question, through the Question Analyser, Search Engine, Paragraph Sorter, to the Answer Processor, and finally to the Human with the answer.]
• Jeopardy Challenge, a compelling and notable way to measure the technology of automatic Question Answering

Slide adapted from David Ferucci
Question Classification

• The process by which a system analyzes a question and labels the question with a given answer type

  Q: “Who was the first Prime Minister of Canada?”

  Answer Type: Person

• It is an integral part of any QA system
  – limits the range of data/corpora to be searched
  – makes the QA problem tractable

Task Definition

• Given a set of questions, build an automated system that can learn the mapping between the questions and the answer types
Task Definition

• Given a set of questions, build an automated system that can learn the mapping between the questions and the answer types

• What’s in a Question?
  – Question Type
  – Answer Type
  – Question Focus
  – Question Topic

Question Type

• Def: An idiomatic categorization of questions for the purpose of distinguishing between different processing strategies and/or answer formats.

  – FACTOID: “How far is it from Earth to Mars?”
  – LIST: “List the names of chewing gums”
  – DEFINITION: “Who is Vlad the Impaler?”
  – YES-NO: “Is Saddam Hussein alive?”
  – OPINION: “What do most Americans think of gun control?”
  – ...

Trivia-like questions where short, factual answers are expected
Answer Type

- **Def:** The class of object (type of sentence) sought by the question.
  - PERSON (from “Who ...”)
  - PLACE (from “Where ...”)
  - DATE (from “When ...”)
  - NUMBER (from “How many ...”)

  but also

  - EXPLANATION (from “Why ...”)
  - METHOD (from “How ...”)

Question Focus

- **Def:** The property or entity that is being sought by the question.
  - “In what state is the Grand Canyon?”
  - “What is the population of Bulgaria?”
  - “What color is a pomegranate?”
Question Focus

- **Def:** The property or entity that is being sought by the question.
  
  - “In what state is the Grand Canyon?”
  - “What is the population of Bulgaria?”
  - “What color is a pomegranate?”

Question Topic

- **Def:** The object (person, place, ...) or event that the question is about. Note that the question might well be about a property of the topic, which will be the question focus.
  
  - “What is the height of Mt. Everest?”
    - **height** is the focus
    - **Mt. Everest** is the topic
Data Creation

- Who creates the Answer Types?
- How and why are certain Answer Types selected?
  - Answer Types are usually tied to the classes recognized by output of Named Entity Recognition Systems
  - Answer Types depends on the final application

TREC QA Data Sets

- Each year a new set of questions is created
- The questions are not labeled with expected answer types
- The named entity tags are generated by BBN’s Identifinder named entity tagger
- The repository contains 31 Answer Types
Examples of Answer Types

<table>
<thead>
<tr>
<th>Example question</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>What do you call a group of geese?</td>
<td>Animal</td>
</tr>
<tr>
<td>Who was Monet?</td>
<td>Biography</td>
</tr>
<tr>
<td>How many types of lemurds are there?</td>
<td>Cardinal</td>
</tr>
<tr>
<td>What is the effect of acid rain?</td>
<td>Cause / Effect / Influence</td>
</tr>
<tr>
<td>What is the street address of the White House?</td>
<td>Contact Info</td>
</tr>
<tr>
<td>Boxing Day is celebrated on what day?</td>
<td>Date</td>
</tr>
<tr>
<td>What is sale?</td>
<td>Definition</td>
</tr>
<tr>
<td>What is another name for nearsightedness?</td>
<td>Disease</td>
</tr>
<tr>
<td>What was the famous battle in 1836 between Texas and Mexico?</td>
<td>Event</td>
</tr>
<tr>
<td>What is the tallest building in Japan?</td>
<td>Facility</td>
</tr>
<tr>
<td>What type of bridge is the Golden Gate Bridge?</td>
<td>Facility Description</td>
</tr>
<tr>
<td>What is the most popular sport in Japan?</td>
<td>Game</td>
</tr>
<tr>
<td>What is the capital of Sri Lanka?</td>
<td>Geo-Political Entity</td>
</tr>
<tr>
<td>Name a Gaelic language.</td>
<td>Language</td>
</tr>
<tr>
<td>What is the world's highest peak?</td>
<td>Location</td>
</tr>
<tr>
<td>How much money does the Sultan of Brunei have?</td>
<td>Money</td>
</tr>
</tbody>
</table>

Distribution of Answer Types

- date, definition, GPE, person ~57%
- facility description, time are seen only 1
N-gram Features

- N-grams are ordered arrangement of \( n \) words
  - \( n=1 \) unigram
  - \( n=2 \) bigram
  - ...

Ex. “How far is it from Earth to Mars?”
  unigrams: \{how, far, is, it\}
  bigrams: \{how far, far is, is it ...\}

*Is it good to use higher order n-grams? Why?*
Question Word Feature

- Use the question word itself as a separate feature

Ex. “How far is it from Earth to Mars?”
   unigrams: {how, far, is, it}
   question word: {how}

But isn’t this very repetitive? Would it actually help or make a difference?

Ex. “What color does litmus paper turn when it comes into contact with a strong acid?”
   avoid misclassifying
   the question as time

Limitations of the Discussed Features

- They will incorrectly classify a question that has multiple types

Ex. **What** question
   - What is the proper name for a female *walrus*?  
     *Type*  
     animal
   - What is Nicholas Cage’s *profession*?
     bio
   - What is the *population* of Seattle?
     cardinal
   - What *caused* the Lynmouth floods?
     cause/effect
   - What is the *telephone number* for the University of Kentucky?
     contact info
   - What is another name for *nearsightedness*?
     disease
Semantic Features

- Let’s incorporate some semantics
- Where does semantic information come from?

- What is the proper name for a female *walrus*?

- What is the *population* of Seattle?

- What is the tallest *mountain*?

- What is the tallest *building* in Japan?

WordNet Features

**Headword Extraction Procedure**

For each question:

- run POS-tagger
- find the first noun phrase and assume it is the main one
- extract the rightmost word tagged as noun
- use WordNet to determine all hypernyms of the headword

*What is the tallest *building* in Japan?*
Web-based Features

- For each question and its headword, submit to the Web lexico-syntactic patterns of the form:

  "* such as headword"

- Retrieve and rerank all semantic classes learned from the Web

  What is the proper name for a female walrus?

Other Features

- Part-of-speech tags
- Noun phrases
- Question length
- Named Entities (helps generalize the question)

  Who is Duke Ellington?  <->  Who is <person_name>?
  Who is Charles Lindbergh?  <->  Who is <person_name>?
Machine Learning Classifier

- The most robust performance is achieved with Support Vector Machines (SVM)

- Training:
  - For each question word (what, how, when, ...) build different SVM classifiers

- Testing:
  - For each unseen question, extract the question word
  - Build a feature vector using the same features which were used in training
  - Use the SVM classifier corresponding to the question word

What is another good measure for the evaluation of question classification?
Evaluation Measures

- Mean Reciprocal Rank (MRR) is used to evaluate a list of possible responses to a query, ordered by probability of correctness.

\[
MRR = \frac{1}{Q} \sum_{i=1}^{Q} \frac{1}{R_{AT}}
\]

- \(Q\) – total number of questions to evaluate
- \(R_{AT}\) – rank at which answer type is found

<table>
<thead>
<tr>
<th>Question</th>
<th>Result of Predicted Answer Type</th>
<th>Correct Answer Type</th>
<th>Rank</th>
<th>Reciprocal Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is Nicholas Cage’s profession?</td>
<td>person, bio, contact info, definition</td>
<td>bio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What is the tallest building in Japan?</td>
<td>entity, location, gpe, facility</td>
<td>facility</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MRR = 0.375
Question Classification Performance

- Baseline model
  - bag-of-word features (i.e. the individual words that make up the question)

- Obtained results with described features

<table>
<thead>
<tr>
<th>Rank R</th>
<th>Baseline</th>
<th>Syntax (bigram)</th>
<th>Semantic</th>
<th>Syntax+Semantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>77.59</td>
<td>78.81</td>
<td>80.39</td>
<td>81.25</td>
</tr>
<tr>
<td>2</td>
<td>86.42</td>
<td>86.85</td>
<td>90.45</td>
<td>88.58</td>
</tr>
<tr>
<td>3</td>
<td>89.51</td>
<td>90.30</td>
<td>93.10</td>
<td>93.32</td>
</tr>
<tr>
<td>4</td>
<td>91.16</td>
<td>91.74</td>
<td>95.55</td>
<td>94.40</td>
</tr>
<tr>
<td>5</td>
<td>92.60</td>
<td>92.82</td>
<td>97.34</td>
<td>95.33</td>
</tr>
<tr>
<td>10</td>
<td>95.33</td>
<td>95.40</td>
<td>95.40</td>
<td>97.20</td>
</tr>
<tr>
<td>MRR</td>
<td>.8437</td>
<td>.8517</td>
<td>.8727</td>
<td>.8737 (+3)</td>
</tr>
</tbody>
</table>

Error Analysis

- Inconsistency in the labeled data

Who is *Duke Ellington*?
Who is *Charles Lindbergh*?

What does *CNN* stand for?
What does *USPS* stand for?
Error Analysis

• Ambiguous answer types

Where was Lincoln assassinated?

<table>
<thead>
<tr>
<th>Type</th>
<th>facility</th>
<th>location</th>
</tr>
</thead>
</table>

Where was the actress, Marion Davies, buried?

<table>
<thead>
<tr>
<th>Type</th>
<th>facility</th>
<th>location</th>
</tr>
</thead>
</table>

Error Analysis

• Inherently difficult questions

What is the name of the Lion King’s son in the movie “The Lion King”?  

<table>
<thead>
<tr>
<th>Type</th>
<th>animal</th>
</tr>
</thead>
</table>

Who developed potlatch?

<table>
<thead>
<tr>
<th>Type</th>
<th>nationality</th>
</tr>
</thead>
</table>
Error Analysis

• POS-tagging misclassifications

What US Government agency registers trademarks?

What/ WP US/NNP Government/NN agency/NN registers/NNS trademarks/NNS ?/. 

Possible solution: do a pull parse tree of the question

Error Analysis

• WordNet coverage and ambiguity

What cereal goes “snap, crackle, pop”? 

WordNet Sense 1

- cereal, cereal grass
  => grass
  => gramineous plant, graminaceous plant
  => herb, herbaceous plant
  => vascular plant
Can you build a QC system using graphs?

In what city was Peter born?
What is Peter’s birthplace?
What is the birthplace of Peter?
Name Peter’s birthplace.

What is the tallest mountain?
What is the highest mountain in the world?
Name the highest mountain.
What is the name of the tallest mountain?