Question Answering with Knowledge Bases, Web and Beyond

Scott Wen-tau Yih & Hao Ma
Search Engine Evolves

San Diego - San Diego Hotels | Things To Do, Activities, ...
www.sandiego.com
SanDiego.com is the best source for all your San Diego vacation needs from deals on hotels and attractions to exciting nightlife and fun things to do around town.

Things To Do
Find the best San Diego things to do and tours of Southern California ...

Hotels
Browse the top San Diego hotels and find the right accommodations to ...

Restaurants
San Diego Restaurants.
SanDiego.com's guide to San ...

See results only from sandiego.com

Attractions
San Diego Attractions. San Diego attractions range from the exciting ...

Best of San Diego
Plan your trip with the Best of San Diego travel guide, featuring the ...

Theme Parks
San Diego theme parks range from Knott's Soak City to SeaWorld. ...

San Diego - Official Site
https://www.sandiego.gov
With its great weather, miles of sandy beaches, and major attractions, San Diego is known worldwide as one of the best tourist destinations and a great place for ...

The Official Travel Resource for the San Diego Region
www.sandiego.org
Find Information on San Diego hotels, restaurants and events for visitors, meeting planners and travel agents.

The Official Travel Resource for the San Diego Region
www.sandiego.org
Find Information on San Diego hotels, restaurants and events for visitors, meeting planners and travel agents.

San Diego - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/San_Diego
San Diego / san di e ˈ goʊr (Spanish for "Saint Didacus") is a major city in California, on the coast of the Pacific Ocean in Southern California, ...

Climates - San Diego County, California - List of people from San Diego - Balboa Park

San Diego - San Diego Hotels | Things To Do, Activities, Tours
www.sandiego.com
SanDiego.com is the best source for all your San Diego vacation needs from deals on hotels and attractions to exciting nightlife and fun things to do around ...

Things to do in San Diego - Best of San Diego - San Diego Attractions - Theme Parks

City of San Diego Official Website
https://www.sandiego.gov
Reference for official information about the city. Specifically in the areas of city and local government.

Things to do in San Diego, California | Facebook

University of California, San Diego
https://uclal.edu
The University California, San Diego is one of the world's leading public research universities, located in beautiful La Jolla, California.
Search Engine Evolves

San Diego
City
San Diego is a major city in California, on the coast of the Pacific Ocean in Southern California, approximately 120 miles south of Los Angeles and immediately adjacent to the border with Mexico.

Weather

- 63 °F Mostly Cloudy
- H 63 °F · L 63 °F

Webcams

- La Jolla, Windansarsa Beach Cam
- SanDiego Cam
- Elephant Cam

Points of interest

- Balboa Park
- San Diego Zoo
- Mission San Diego de Alcalá
- SeaWorld San Diego
- San Diego Zoo Safari Park

People also search for

- Los Angeles
- San Francisco
- San Jose
- Phoenix
- Seattle

Explore more

Largest cities by population in California
Search Engine Evolves

places to go in san diego

San Diego - Points of interest

Balboa Park, San Diego Zoo, Mission San Diego de Alcalá, SeaWorld San Diego, San Diego Zoo Safari Park, Old Town San Diego State Historic Park, Westfield Horton Plaza, Hotel del Coronado, Point Loma, San Diego
Question and Answering in Modern Search Engines
Question and Answering in Modern Search Engines

Search Results:
- Tom Hanks Movies with Meg Ryan

Also try: Joe Versus the Volcano · Tom Hanks Meg Ryan Movies Together · All ...

Movies of Tom Hanks starring Meg Ryan
- Sleepless in Seattle (1993) ★★★★★
- You've Got Mail (1998) ★★★★★
- Joe Versus the Volcano (1990) ★★★★★
- Hope for Haiti Now: A Global Benefit for E...
First movie of Tom Hanks starring Meg Ryan

Joe Versus the Volcano (1990)
Director of first movie of Tom Hanks starring Meg Ryan

John Patrick Shanley

Joe Versus the Volcano (1990) - IMDb
www.imdb.com/title/tt0099802
Rating: 5.7/10 · 25,640 ratings · Comedy/Romance · PG · 102 min
Joe Versus the Volcano PG ... Director: John Patrick Shanley. Writer: John Patrick Shanley. Stars: Tom Hanks, Meg Ryan, Lloyd Bridges | See full cast and crew »

Meg Ryan Reteams With Tom Hanks for Ithaca, Actress Set ... www.eonline.com/news/505216/meg-ryan-reteams-with-tom-hanks-for... Jan 29, 2014 · Meg Ryan and Tom Hanks are teaming ... latest to step into the role of director. ... and was instrumental in making the first film such a ...
What is the largest animal in history?

A member of the order Cetacea, the blue whale (Balaenoptera musculus), is believed to be the largest animal ever to have lived.

Largest organisms - Wikipedia, the free encyclopedia

[en.wikipedia.org/wiki/Largest_animal](en.wikipedia.org/wiki/Largest_animal)

What is the longest river in the world?

The Nile

The Nile in Africa has long been considered the world’s longest river, but there is some debate about the definition of the length of a river that leads some to claim that the Amazon in South America is longer. The claim that the Amazon is longer is reached by measuring the river plus the adjacent Pará estuary and the longest connecting tidal canal. The approximate length of the rivers with the debated measurements are:

References:
- [en.wikipedia.org/wiki/List_of_rivers_by_length](en.wikipedia.org/wiki/List_of_rivers_by_length)
- [en.wikipedia.org/wiki/Amazon_River#Dispute_regarding_length](en.wikipedia.org/wiki/Amazon_River#Dispute_regarding_length)
Question and Answering in Modern Search Engines

---

**Main NLP/CL 2016 Conference Deadlines**

<table>
<thead>
<tr>
<th>Conference</th>
<th>Submission Date</th>
<th>Notification Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACL 2016 (long papers)</td>
<td>Mar 18</td>
<td>May 24</td>
</tr>
<tr>
<td>Interspeech 2016</td>
<td>Mar 23</td>
<td>Jun 10</td>
</tr>
<tr>
<td>CoNLL 2016</td>
<td>May 06</td>
<td>Jun 01</td>
</tr>
</tbody>
</table>

12 more rows, 2 more columns

---

**(The Old) NLP/CL Conference Calendar**

[https://www.cs.rochester.edu/~tetrau/confenices.html](https://www.cs.rochester.edu/~tetrau/confenices.html)  
University of Rochester
Conversational Question Answering

Tom Cruise

Tom Cruise is an American actor and filmmaker. Cruise has been nominated for three Academy Awards and has won three Golden Globe Awards. He started his career at age...
How tall is Katie Holmes?
Conversational Question Answering

How about Nicole Kidman's height? I found this for you.

Nicole Kidman · Height

5 feet 11 inches
(1.80 meter)

See more about Nicole Kidman →

Nico (given name) - Wikipedia, the free encyclopedia
https://en.m.wikipedia.org/wiki/Nicole_(given_name)
Where was she born

Honolulu, HI
Medal Count by sport for france and china as bar chart sorted by country

Show medal count; sport; and areas that medalled in sport where area is france or china as stacked bar chart
which player scored the most unassisted goals per world cup

Show players that scored goals and world cups, where assist player name is N/A sorted by number of goals descending

Count of Goals by Player Name, and WorldCupName

- David Villa
- Gabriel Batistuta
- Fernando Hierro
- Hristo Stoichkov
- David Beckham
- Diego Forlan
- Gary Lineker
- Lothar Matthaus
Natural Language Understanding

- **Question-answering machine** [Simmons CACM-65]
  - General-purpose language processors that communicate with users in natural language (e.g., English)
  - Deal with statements and/or questions

http://csunplugged.org/turing-test
Categories of (Early) QA Systems

• **List-structured database systems**
  • Organizing knowledge (e.g., kinship) in list DB

• **Graphic database systems**
  • Map text and graphic data (e.g., pictures, diagrams) to the same logical representations

• **Text-based systems**
  • Matching questions and text in a corpus to find answers

• **Logical inference systems**
  • Textual entailment, answering science text book questions & algebra word problems
Baseball [Green, Wolf, Chomsky & Laughery, 1961]

• How many games did the Yankees play in July?

• Step 1: Simple dictionary-based syntactic analysis
  • (How many games) did (the Yankees) play (in (July))?  

• Step 2: Semantic analysis that builds “spec”
  • “Who” → (“team” = ?)  
  • Conditions (e.g., “winning”, “how many”) → routines

• Step 3: Execution

Example taken from [Simmons, 1965]
The **Picture Language Machine** [Krisch, 1964]

- *Is the statement true?*
  
  *All circles are black circles.*

- Both pictures and text are translated into logical language
  - Circle(a), Black(a), Bigger(a, b), Between(a, b, c)
  - $(\forall x)[\text{Circle}(x) \supset (\exists y)[\text{Circle}(y) \land \text{Black}(y) \land (x = y)]]$
Protosynthex [Simmons+, 1964]
Answer Questions from an Encyclopedia

- Matching questions & text in dependency logic [Hays 1962]

Q: What do worms eat?

A1: Worms eat grass

A2: Horses with worms eat grain

Complete Agreement

Partial Agreement
Student [Bobrow 1964]

• The first algebra problem solver
  • Translate a set of English statements to mathematical equations

• Step 1: Simplify text and annotate operators
  • “twice” → “two times”, “the square of” → “square”
  • Tag operators like “plus”, “percent”, “times”

• Step 2: Heuristics to break problem into simple sentences

• Step 3: Mapping sentences to equations
  • Rules based on dictionary of words and numbers
Lessons from Old QA Systems

• **Limited success**
  • Small & limited domains and scopes
    • Often work only on well-controlled, specialized subset of English
  • Not data-driven (e.g., machine learning approaches)
    • Mostly rule-based, potentially brittle
    • Lacks rigorous evaluation

• **Open questions** [Simmons 1965]
  • Meaning representation & the need of formal languages
  • Syntactic and semantic disambiguation
  • Combine partial answers from various sources
Categories of Modern QA Systems/Problems

• Factoid questions
  • Informational queries about facts of entities
  • Competitions (Jeopardy! & Quiz Bowl)

• Narrative questions
  • Opinion, instructions (how-to questions)

• Multi-modal
  • Visual QA
  • Travel Assistant

• AI ability tests
  • Reading comprehension
  • Elementary School Science and Math Tests
Factoid Questions

**When did Minnesota become a state?**

May 11, 1858
Minnesota · Founded

**Who was Katy Perry's husband?**

Russell Brand
(2010 - 2012)
Katy Perry · Spouse

**When was Washington founded?**

November 11, 1889
Washington, DC History | washington.org
Mobile-friendly · Founded on July 16, 1790, Washington DC is unique among American cities because it was established by the
Visual Question Answering [Agrawal et al.]

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

Is this person expecting company?
What is just under the tree?

Does it appear to be rainy?
Does this person have 20/20 vision?
James the Turtle was always getting in trouble. Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.

One day, James thought he would go into town and see what kind of trouble he could get into. He went to the grocery store and pulled all the pudding off the shelves and ate two jars. Then he walked to the fast food restaurant and ordered 15 bags of fries. He didn't pay, and instead headed home.

His aunt was waiting for him in his room. She told James that she loved him, but he would have to start acting like a well-behaved turtle.

After about a month, and after getting into lots of trouble, James finally made up his mind to be a better turtle.

1) What is the name of the trouble making turtle?
   A) Fries
   B) Pudding
   C) James
   D) Jane

2) What did James pull off of the shelves in the grocery store?
   A) pudding
   B) fries
   C) food
   D) splinters
Data Sources

- Structured data
  - Databases & Knowledge bases
- Semi-structured data
  - Web tables
- Unstructured text
  - Newswire corpora
  - Web
Paradigms

- **Semantic parsing**
  - Answer questions using knowledge bases
- **Information Retrieval**
  - Text matching
- **Human intelligence**
  - Community QA
  - Social QA ([I’m an Expert](#)) [Richardson & White, WWW-2011]
General Technological Challenges

- Question analysis
  - Answer type
  - Slot filling
  - Semantic parsing
- Text/Data analysis
- Paraphrasing & Matching
  - Handle variations of questions
  - Ontology matching
- Search complexity
Roadmap

• Question Answering with Knowledge Bases
  • Introduction to modern large-scale knowledge bases
  • Datasets and state-of-the-art approaches

• Question Answering with the Web
  • Problem setting and general system architecture
  • Essential natural language analysis
  • Leveraging additional information sources

• Question Answering for Testing Machine Intelligence
  • Reading comprehension
  • Reasoning questions
Question Answering with Knowledge Bases
Answer Questions Using Structured Data

• General problem setting
  • Information Source: A “database”
    • Collections of records
    • Tables
    • Large-scale DB with complex schema
  • Input: A natural language question (instead of a formal “query”)
  • Output: Answer
Baseball [Green, Wolf, Chomsky & Laughery, 1961]

• How many games did the Yankees play in July?

Example taken from [Simmons, 1965]
LUNAR [Woods, 1973]

- Give me all lunar samples with Magnetite.
- How many samples contain Titanium?
Geoquery [Zelle & Mooney, 1996]

- What is the capital of the state with the largest population?
- What are the major cities in Kansas?

<table>
<thead>
<tr>
<th>Type</th>
<th>Form</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>country</td>
<td>countryid(Name)</td>
<td>countryid(usa)</td>
</tr>
<tr>
<td>city</td>
<td>cityid(Name, State)</td>
<td>cityid(austin,tex)</td>
</tr>
<tr>
<td>state</td>
<td>stateid(Name)</td>
<td>stateid(texas)</td>
</tr>
<tr>
<td>river</td>
<td>riverid(Name)</td>
<td>riverid(colorado)</td>
</tr>
<tr>
<td>place</td>
<td>placeid(Name)</td>
<td>placeid(pacific)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Form</th>
<th>Predicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>capital(C)</td>
<td>is a capital (city).</td>
</tr>
<tr>
<td>city(C)</td>
<td>is a city.</td>
</tr>
<tr>
<td>major(X)</td>
<td>X is major.</td>
</tr>
<tr>
<td>place(P)</td>
<td>P is a place.</td>
</tr>
<tr>
<td>river(R)</td>
<td>R is a river.</td>
</tr>
<tr>
<td>state(S)</td>
<td>S is a state.</td>
</tr>
<tr>
<td>capital(C)</td>
<td>C is a capital (city).</td>
</tr>
<tr>
<td>area(S,A)</td>
<td>The area of S is A.</td>
</tr>
<tr>
<td>capital(S,C)</td>
<td>The capital of S is C.</td>
</tr>
<tr>
<td>equal(V,C)</td>
<td>variable V is ground term C.</td>
</tr>
<tr>
<td>density(S,D)</td>
<td>The (population) density of S is D</td>
</tr>
<tr>
<td>elevation(P,E)</td>
<td>The elevation of P is E.</td>
</tr>
<tr>
<td>high_point(S,P)</td>
<td>The highest point of S is P.</td>
</tr>
<tr>
<td>low_point(S,P)</td>
<td>The lowest point of S is P.</td>
</tr>
</tbody>
</table>

Example taken from [Zelle & Mooney, 1996]
Early Work

- **Small scale & domain-specific KBs**
  - Simple schema
  - Small numbers of entities and relations
  - Limited set of sensible questions

- **Approaches**
  - Ad-hoc methods (e.g., manually crafting rules) can be quite effective
  - Semantic parsing (of questions)

- **Issues**
  - Not clear if the methods are scalable
  - Cannot support “open-domain” question answering
Modern Large-scale Knowledge Bases

- Freebase: 46m entities, 2.6b facts
- Microsoft Satori: 852m entities, 18b facts
Entity-centric
Properties & Relations between Entities

Seattle

- **NFL championships:** 2013
- **Head coach:** Pete Carroll
- **Founded:** 1976
- **Division:** NFC West

- **Address:** 400 Broad St, Seattle, 98109
- **Phone:** (800) 937-9582
- **Opened:** Apr 21, 1962
- **Height:** 605 feet (184.41 m)
- **Floors:** 6

Headquarters

- **Founded:** Mar 30, 1971 · Pike Place Market
- **Customer service:** +1 800-782-7282
- **CEO:** Howard Schultz
- **Founders:** Jerry Baldwin · Zev Siegl · Gordon Bowker

Location

Population: 652,405 (2013)
Area: 142.55 sq miles (369.20 km²)
Mayor: Ed Murray

Pike Place Market

CEO: Howard Schultz
Founders: Jerry Baldwin · Zev Siegl · Gordon Bowker
Subject-Predicate-Object Triples in Freebase

Seattle Seahawks

Pete Carroll

m.070xg, american_football/football_team/current_head_coach, m.02ttv2
Representing Multi-argument Relations

- Seattle Seahawks – sports.sports_team.roster

<table>
<thead>
<tr>
<th>Player</th>
<th>Number</th>
<th>Position</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russell Wilson</td>
<td>3</td>
<td>Quarterback</td>
<td>2012</td>
<td>-</td>
</tr>
<tr>
<td>Alan Branch</td>
<td>99</td>
<td>Defensive tackle</td>
<td>2011</td>
<td>2012</td>
</tr>
<tr>
<td>Marshawn Lynch</td>
<td>24</td>
<td>Running back</td>
<td>2010</td>
<td>2016</td>
</tr>
<tr>
<td>Richard Sherman</td>
<td>25</td>
<td>Cornerback</td>
<td>2011</td>
<td>-</td>
</tr>
</tbody>
</table>

...
Representing Multi-argument Relations

- Seattle Seahawks – sports.sports_team.roster

<table>
<thead>
<tr>
<th>Player</th>
<th>Number</th>
<th>Position</th>
<th>From</th>
<th>To</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russell Wilson</td>
<td>3</td>
<td>Quarterback</td>
<td>2012</td>
<td>-</td>
</tr>
<tr>
<td>Alan Branch</td>
<td>99</td>
<td>Defensive tackle</td>
<td>2011</td>
<td>2012</td>
</tr>
<tr>
<td>Marshawn Lynch</td>
<td>24</td>
<td>Running back</td>
<td>2010</td>
<td>2016</td>
</tr>
<tr>
<td>Richard Sherman</td>
<td>25</td>
<td>Cornerback</td>
<td>2011</td>
<td>-</td>
</tr>
</tbody>
</table>

- Compound Value Type (CVT) Nodes
  - Seattle Seahawks – sports/sports_team/roster – CVT1
  - CVT1 – sports/sports_team_roster/player – Russel Wilson
  - CVT1 – sports/sports_team_roster/number – 3
Question Answering with Knowledge Base

- **Large-scale Knowledge Base**
  - Properties of billions of entities
  - Plus relations among them

- **Question Answering**
  
  "What are the names of Obama’s daughters?"

  \[ \lambda x. \text{parent}(\text{Obama}, x) \land \text{gender}(x, \text{Female}) \]
WebQuestions Dataset [Berant+ 13]

- What character did Natalie Portman play in Star Wars? ⇒ Padme Amidala
- What currency do you use in Costa Rica? ⇒ Costa Rican colon
- What did Obama study in school? ⇒ political science
- What do Michelle Obama do for a living? ⇒ writer, lawyer
- What killed Sammy Davis Jr? ⇒ throat cancer

5,810 questions crawled from Google Suggest API and answered using Amazon MTurk
- 3,778 training, 2,032 testing
- A question may have multiple answers ⇒ using Avg. F1 (≈ accuracy)
Approaches

- Semantic Parsing
  - Generic semantic parsing and then ontology matching
  - KB-specific semantic parsing
- Information Extraction
- Embedding
Generic Semantic Parsing (e.g., [Kwiatkowski+ 13])

Who is Justin Bieber’s sister?

Jazmyn Bieber

Knowledge Base

\[ \lambda x. \text{sister	extunderscore of}(\text{justin	extunderscore bieber}, x) \]

\[ \lambda x. \text{sibling	extunderscore of}(\text{justin	extunderscore bieber}, x) \land \text{gender}(x, \text{female}) \]
KB-Specific Semantic Parsing (e.g., [Berant+ 13])

Who is Justin Bieber’s sister?

Jazmyn Bieber

\[ \lambda x. \text{sibling\_of}(\text{justin\_bieber}, x) \land \text{gender}(x, \text{female}) \]
Key Challenges

• Language mismatch
  • Lots of ways to ask the same question
    “Who played the role of Meg on Family Guy?”
    “What is the name of the actress for Meg on Family Guy?”
    “In the TV show Family Guy, who is the voice for Meg?”

• Need to map questions to the predicates defined in KB
  tv.tv_program.regular_cast – tv.regular_tv_appearance.actor

• Large search space
  • Some Freebase entities have >160,000 immediate neighbors

• Compositionality
  • “What movies are directed by the person who won the most Academy and Golden Globe awards combined?”
**SEMPRE – \( \lambda \)-DCS** [Liang, 2013]

- \( \lambda \)-DCS (lambda dependency-based compositional semantics)
  - Utterance: “people who have lived in Seattle”
  - Logical form (lambda calculus): \( \lambda x. \exists e. \text{PlacesLived}(x, e) \land \text{Location}(e, \text{Seattle}) \)
  - Logical form (lambda DCS): \( \text{PlacesLived}.\text{Location}.\text{Seattle} \)

- Unary: Seattle \( \lambda x. [x = \text{Seattle}] \)
- Binary: PlaceOfBirth \( \lambda x. \lambda y. \text{PlaceOfBirth}(x, y) \)
- Join: “people born in Seattle” PlaceOfBirth.Seattle \( \lambda x. \text{PlaceOfBirth}(x, \text{Seattle}) \)
- Intersection: “scientists born in Seattle” Profession.Scientist \( \sqcap \) PlaceOfBirth.Seattle \( \lambda x. \text{Profession}(x, \text{Scientist}) \land \text{PlaceOfBirth}(x, \text{Seattle}) \)
Bridging: Hypothesizing predicates to be connected when the type constraints are satisfied

“What government does Chile have?”
“What actors are in Top Gun?”
“What is Italy money?”

Type.FormOfGovernment
Chile
What political party founded by Henry Clay?  ...  What event involved the people Henry Clay?

\[
\text{Type.PoliticalParty} \sqcap \text{Founder.HenryClay}  \quad ...  \quad \text{Type.Event} \sqcap \text{Involved.HenryClay}
\]
Austin is the capital of Texas. What is the capital of Texas?

- **Word Nodes (Ovals)**
  - word nodes are connected via syntactic dependencies
- **Entity Nodes (Rectangles)**
- **Mediator Nodes (Circles)**
  - Represent events
  - Binary predicates
- **Type nodes (Rounded rectangles)**
  - Unary predicates
- **Math nodes (Diamonds)**
  - e.g., Aggregation Functions

*Fig. 2 of [Reddy et al., 2014]*
Each span is a mapping of a single-relation question:

**Question Pattern:**
“Who is the director of Forrest Gump?”
(Forrest Gump, Director, ?)

- Patterns from mining Bing query logs.

* Few questions in WebQuestions are with a long chain like this.

_“Machine Translation” [Bao et al., ACL-14]_

---

**Fig.1 of [Bao et al., 2014]**
Staged Query Graph Generation [Yih et al. ACL-15]

Core idea

- Proposing a new semantic parse language — query graph
  - Resembles subgraphs of the knowledge base
  - Can be directly mapped to an executable query (e.g., SQL, SPARQL)

- Reducing semantic parsing to a search problem
  - Grows the candidate query graph through staged state-actions
Who first voiced Meg on Family Guy?

\[ \lambda x. \exists y. \text{cast}(\text{FamilyGuy}, y) \land \text{actor}(y, x) \land \text{character}(y, \text{MegGriffin}) \]
Query Graph – Topic Entity

Who first voiced Meg on Family Guy?
Link Topic Entity

- An advanced entity linking system for short text

- Prepare surface-form lexicon $\mathcal{L}$ for entities in the KB
- Entity mention candidates: all consecutive word sequences in $\mathcal{L}$, scored by the statistical model
- Up to 10 top-ranked entities are considered as topic entity
**Query Graph – Core Inferential Chain**

Who first voiced Meg on *Family Guy*?

\{'cast-actor, producer, awards_won-winner\}
Identify Core Inferential Chain

- Relationship between topic and answer ($x$) entities
- Explore two types of paths
  - Length 1 to non-CVT node
  - Length 2 where $y$ can be grounded to CVT

Who first voiced Meg on **Family Guy**?

{cast-actor, writer-start, genre}
Relation Matching using Deep Convolutional Neural Networks (DSSM [Shen+ 14])

- Input is mapped to two $k$-dimensional vectors
- Probability is determined by softmax of their cosine similarity

$$P(R|P) = \frac{\exp(\cos(y_R, y_P))}{\sum_{R'} \exp(\cos(y_{R'}, y_P))}$$

who voiced meg on \texttt{<e>} \texttt{cast-actor}
Query Graph - Constraints

Who first voiced Meg on Family Guy?
Augment Constraints

• Who first voiced Meg on Family Guy?

\[ \lambda x. \exists y. \text{cast}(\text{FamilyGuy}, y) \land \text{actor}(y, x) \]

• One or more constraint nodes can be added to \( y \) or \( x \):
  • \( y \): Additional property of this event (e.g., \( \text{character}(y, \text{MegGriffin}) \))
  • \( x \): Additional property of the answer entity (e.g., \( \text{gender} \))

• Only subset of constraint nodes are considered
  • e.g., entities detected in the question
Learning Reward Function $\gamma$

- Judge whether a query graph is a correct semantic parse
- Log-linear model with pairwise ranking objective [Burges 10]

Who first voiced Meg on Family Guy?
Learning Reward Function $\gamma$

- Judge whether a query graph is a correct semantic parse
- Log-linear model with pairwise ranking objective [Burges 10]

Who first voiced Meg on Family Guy?

\[ \gamma(s_7) \quad \text{from} \quad \text{character} \quad \text{Meg Griffin} \quad \text{cast} \quad \text{actor} \quad \text{Family Guy} \]

\[ \gamma(s_3) \]

\[ \gamma(s_3) > \gamma(s_7) \]
Learning Reward Function – Features

$q = \text{Who first voiced Meg on Family Guy?}$

\[ s = \]

- **Topic Entity**
  - Entity linking scores

- **Core Inferential Chain**
  - Relation matching scores (NN models)

- **Constraints: Keyword and entity matching**
  - $\text{ConstraintEntityWord}("\text{Meg Griffin}", q) = 0.5$
  - $\text{ConstraintEntityInQuestion}("\text{Meg Griffin}", q) = 1$

- **Overall**
  - $\text{NumNodes}(s) = 5$
  - $\text{NumAnswers}(s) = 1$
Creating Training Data from Q/A Pairs

Relation Matching (Identifying Core Inferential Chain)

- List all the length 1 & 2 paths from any potential topic entity
- Treat any inferential chain resulting in $F_1 \geq 0.5$ to create positive pairs

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Inferential Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>what was &lt;e&gt; known for</td>
<td>people.person.profession</td>
</tr>
<tr>
<td>what kind of government does &lt;e&gt; have</td>
<td>location.country.form_of_government</td>
</tr>
<tr>
<td>what year were the &lt;e&gt; established</td>
<td>sports.sports_team.founded</td>
</tr>
<tr>
<td>what city was &lt;e&gt; born in</td>
<td>people.person.place_of_birth</td>
</tr>
<tr>
<td>what did &lt;e&gt; die from</td>
<td>people.deceased_person.cause_of_death</td>
</tr>
<tr>
<td>who married &lt;e&gt;</td>
<td>people.person.spouse_s</td>
</tr>
<tr>
<td></td>
<td>people.marriage.spouse</td>
</tr>
</tbody>
</table>
Creating Training Data from Q/A Pairs

Reward Function $\gamma$

- Apply the same best-first search procedure to training data
- Use the $F_1$ score of the query graph as the reward function
- For each question, create 4,000 candidate query graphs
  - All positive ($F_1 > 0$) examples
  - Randomly selected negative examples
Staged Query Graph Generation
Addresses Key Challenges

• Language mismatch
  • Advanced entity linking [Yang & Chang, ACL-15]
  • Relation matching via deep convolutional NN [Shen et al., CIKM-14]

• Large search space
  • Representation power of a parse controlled by staged search actions
  • Grounding partially the question during search

• Compositionality
  • Possible combinations limited by local subgraphs
Information Extraction [Yao & Van Durme, ACL-2014]

• “What is the name of Justin Bieber brother?”

Create lots of features; learn an “answer” classifier (L1-regularized LR)
Embeddings [Bordes et al., EMNLP-2014]

Score $S(q,a)$: How the candidate answer fits the question

Embedding model

Embedding of the question $f(q)$

Embedding matrix $W$

Dot product

Embedding of the subgraph $g(a)$

Binary encoding of the question $\Phi(q)$

Question $q$

“Who did Clooney marry in 1987?”

Freebase subgraph

Detection of Freebase entity in the question

K. Preston

G. Clooney

Honolulu

1987

J. Travolta

Subgraph of a candidate answer $a$ (here K. Preston)

Model

Fig. 1 of [Bordes et al., 2014]
Avg. F1 (Accuracy) on WebQuestions Test Set

Yao-14 | 33.0
Berant-13 | 35.7
Bao-14 | 37.5
Bordes-14b | 39.2
Berant-14 | 39.9
Yang-14 | 41.3
Yih-15 | 52.5
Other Datasets

- **Free917** [Cai & Yates, ACL-13]
  - 917 English questions labeled with lambda expressions with predicates & constants defined in Freebase

- **Simple Questions** [Bordes et al., arXiv:1506.02075]
  - 108,442 questions paired with Freebase triples
  - Multi-argument relations (CVT) don’t seem to be included

- **WebQuestionsSP** ([http://aka.ms/WebQSP](http://aka.ms/WebQSP)) [Yih et al., ACL-16]
  - Full semantic parses of WebQuestions in SPARQL, along with updated answers and additional entity/relation information
Summary

• Recent work on question answering with KB
  • Task: Answering WebQuestions using Freebase
  • Most approaches aim for semantic parsing of questions

• Challenges
  • How to leverage multiple resources to handle language mismatch?
  • How to handle compositionality correctly and efficiently?

• Very active research problem
  • Many new methods being proposed (e.g., [Berant & Liang, TACL-15], [Reddy et al., TACL-16], [Xu et al., ACL-16])
Discussion

• Why is WebQuestions so successful?
  • “Largest” dataset for evaluating semantic parsing
  • A new direction for open-domain question answering

• Is semantic parsing the right approach for QA?
  • Not many alternatives when the information is stored in the DB
  • The derivation of answers is more interpretable; easier to debug
  • Not necessarily the best approach for factoid question answering
Question and Answering with the Web
Who first landed on the Moon?

SELECT ?p
WHERE {?p land-on ?m . ?m target Moon . ?m date ?t .}
ORDER BY ?t LIMIT 1

Issues:
• Semantic parsing is difficult due to ontology mismatch
• Knowledge base is incomplete (missing entities/relations/attribute values)
Knowledge Base is largely incomplete

<table>
<thead>
<tr>
<th>Relation</th>
<th>Percentage unknown</th>
<th>All 3M</th>
<th>Top 100K</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROFESSION</td>
<td>68%</td>
<td>24%</td>
<td></td>
</tr>
<tr>
<td>PLACE OF BIRTH</td>
<td>71%</td>
<td>13%</td>
<td></td>
</tr>
<tr>
<td>NATIONALITY</td>
<td>75%</td>
<td>21%</td>
<td></td>
</tr>
<tr>
<td>EDUCATION</td>
<td>91%</td>
<td>63%</td>
<td></td>
</tr>
<tr>
<td>SPOUSES</td>
<td>92%</td>
<td>68%</td>
<td></td>
</tr>
<tr>
<td>PARENTS</td>
<td>94%</td>
<td>77%</td>
<td></td>
</tr>
<tr>
<td>CHILDREN</td>
<td>94%</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>SIBLINGS</td>
<td>96%</td>
<td>83%</td>
<td></td>
</tr>
<tr>
<td>ETHNICITY</td>
<td>99%</td>
<td>86%</td>
<td></td>
</tr>
</tbody>
</table>

Knowledge Base Completion via Search-Based Question Answering [Robert West, et al., WWW 2014]
Florence Cathedral

The Cattedrale di Santa Maria del Fiore is the main church of Florence, Italy. Il Duomo di Firenze, as it is ordinarily called, was begun in 1296 in the Gothic style ... It remains the largest brick dome ever constructed. en.wikipedia.org
Knowledge Bases

**Issues:**
- Semantic parsing is difficult due to ontology mismatch
- Knowledge base is incomplete (missing entities/relations/attribute values)

**Advantages:**
- Contains abundant information
- Redundancy on the Web could help confirm the answers
Web Question and Answering

- Entity Retrieval/Finding
- Factoid Answer based on Web Documents
- Factoid Answer based on Tables
Entity Retrieval/Finding

Bing search for famous basketball players:
- Michael Jordan
- LeBron James
- Kobe Bryant
- Magic Johnson
- Larry Bird
- Wil Chamberlain (1925 - 1979)
- Kareem Abdul-Jabbar
- Shaquille O'Neal

Bing search for Italian composers:
- Giacomo Puccini (1858 - 1924)
- Gioacchino Rossini (1792 - 1868)
- Ennio Morricone
- Claudio Monteverdi (1567 - 1643)
- Vincenzo Bellini (1801 - 1835)
- Giovanni Pierluigi da Palestrina (1525 - 1594)
- Jean Baptiste Lully (1632 - 1687)
- Nino Rota (1911 - 1979)
Entity Retrieval/Finding

- TREC Entity Track (2009 – 2011)
  - Related Entity Finding Task
  - Given
    - Input entity
    - Type of the target entity (PER/ORG/LOC)
    - Narrative (describing the nature of the relation in free text)
  - Return related entities
Entity Retrieval/Finding

- A typical pipeline

Entity Retrieval/Finding

- Three component model

\[ p(e|E, T, R) \propto p(e|E) \cdot p(T|e) \cdot p(R|E, e) \]

Related Entity Finding Based on Co-Occurrence [Balog, et al., TREC 2009]
Entity Retrieval/Finding

\[ P(R|E,e) = P(R|\Theta_{Ee}) = \prod_{i \in R} P(t|\Theta_{Ee})^{n(t,R)} \]

\[ P(t|\Theta_{Ee}) = \frac{1}{|D_{Ee}|} \sum_{d \in D_{Ee}} P(t|\Theta_{d}) \]

\[ P(t|\Theta_{d}) = \frac{n(t,d) + \mu \cdot P(t)}{\sum_{i} n(t',d) + \mu} \]

Related Entity Finding Based on Co-Occurrence [Balog, et al., TREC 2009]
Entity Retrieval/Finding

Model A

\[ p(e, m = 1 | R, S) \propto p(R | e, S) p(e | S) \sum_{t_R} \sum_{t_e} p(m = 1 | t_e, t_R) p(t_e | e) p(t_R | R) \]

Model B

\[ p(e, m = 1 | R) \propto p(R | e) p(e) \sum_{t_R} \sum_{t_e} p(m = 1 | t_e, t_R) p(t_e | e) p(t_R | R) \]

## Entity Retrieval/Finding

**Input Entity:** Dow Jones  
**Target Entity Type:** Organization

**Narrative:** Find companies that are included in the Dow Jones industrial average

<table>
<thead>
<tr>
<th>$p(m \mid e, R)$</th>
<th>$p(R \mid e)p(e)$</th>
<th>MA</th>
<th>$p(R \mid e, S)p(e \mid S)$</th>
<th>MB</th>
</tr>
</thead>
<tbody>
<tr>
<td>nasdaq</td>
<td>microsoft</td>
<td>boeing</td>
<td>coca cola</td>
<td>boeing</td>
</tr>
<tr>
<td>bloomberg</td>
<td>boeing</td>
<td>ibm</td>
<td>boeing</td>
<td>coca cola</td>
</tr>
<tr>
<td>ibm</td>
<td>federal reserve</td>
<td>pfizer</td>
<td>cnnmoney</td>
<td>microsoft</td>
</tr>
<tr>
<td>news corporation</td>
<td>european</td>
<td>coca cola</td>
<td>futures</td>
<td>nasdaq</td>
</tr>
<tr>
<td>Yahoo</td>
<td>coca cola</td>
<td>intel</td>
<td>microsoft</td>
<td>ibm</td>
</tr>
<tr>
<td>atari</td>
<td>uw</td>
<td>alcoa</td>
<td>pfizer</td>
<td>intel</td>
</tr>
<tr>
<td>washington post</td>
<td>ibm</td>
<td>cnnmoney</td>
<td>alcoa</td>
<td>merck</td>
</tr>
<tr>
<td>boeing</td>
<td>intel</td>
<td>mcdonald’s</td>
<td>ibm</td>
<td>dupont</td>
</tr>
<tr>
<td>stanford</td>
<td>futures</td>
<td>merck</td>
<td>federal reserve</td>
<td>caterpillar</td>
</tr>
<tr>
<td>enterprise media group</td>
<td>merck</td>
<td>microsoft</td>
<td>mcdonald’s</td>
<td>stanford</td>
</tr>
</tbody>
</table>

Entity Retrieval/Finding

- Knowledge base are largely incomplete

<table>
<thead>
<tr>
<th>Relation</th>
<th>Percentage unknown</th>
<th>All 3M</th>
<th>Top 100K</th>
</tr>
</thead>
<tbody>
<tr>
<td>PROFESSION</td>
<td>68%</td>
<td>24%</td>
<td></td>
</tr>
<tr>
<td>PLACE OF BIRTH</td>
<td>71%</td>
<td>13%</td>
<td></td>
</tr>
<tr>
<td>NATIONALITY</td>
<td>75%</td>
<td>21%</td>
<td></td>
</tr>
<tr>
<td>EDUCATION</td>
<td>91%</td>
<td>63%</td>
<td></td>
</tr>
<tr>
<td>SPOUSES</td>
<td>92%</td>
<td>68%</td>
<td></td>
</tr>
<tr>
<td>PARENTS</td>
<td>94%</td>
<td>77%</td>
<td></td>
</tr>
<tr>
<td>CHILDREN</td>
<td>94%</td>
<td>80%</td>
<td></td>
</tr>
<tr>
<td>SIBLINGS</td>
<td>96%</td>
<td>83%</td>
<td></td>
</tr>
<tr>
<td>ETHNICITY</td>
<td>99%</td>
<td>86%</td>
<td></td>
</tr>
</tbody>
</table>

Entity Retrieval/Finding techniques can be used in Knowledge Base Completion

Knowledge Base Completion via Search-Based Question Answering [Robert West, et al., WWW 2014]
Knowledge Base Completion via Search-Based Question Answering [Robert West, et al., WWW 2014]
Entity Retrieval/Finding

• Challenges
  • The TREC’s related entity finding track is relatively easy since the “query intent” is known

  Input Entity: Dow Jones  Target Entity Type: Organization
  Narrative: Find companies that are included in the Dow Jones industrial average

• In real world search engines, we need to understand the intent of queries

  Companies in Dow Jones industrial
Factoid Answer based on Web Documents
Who first landed on the Moon?

Apollo 11 was the spaceflight that landed the first humans on the Moon, Americans Neil Armstrong and Buzz Aldrin, on July 20, 1969, at 20:18 UTC.

Factoid Answer based on Web Documents

- Typical Architecture of Web QnA

Open Domain Question and Answering via Semantic Enrichment [Huan Sun, et al., WWW 2015]
Factoid Answer based on Web Documents

- Detailed Architecture

Question Answering [Dan Jurafsky, Stanford]
Factoid Answer based on Web Documents

- **QUESTION PROCESSING**
  - Detect question type, answer type
  - Formulate queries to send to a search engine

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

- **ANSWER PROCESSING**
  - Extract candidate answers
  - Rank candidates

Question Answering [Dan Jurafsky, Stanford]
Factoid Answer based on Web Documents

- **Answer Type Detection: Name Entities**
  - Who first landed on the moon?
    - Person
  - Where is the headquarter of Microsoft?
    - Location
  - What is the largest country in terms of population?
    - Country
  - Highest flying bird
    - Animal/Bird

Question Answering [Dan Jurafsky, Stanford]
Factoid Answer based on Web Documents

- 6 coarse classes
  - ABBEVIATION, ENTITY, DESCRIPTION, HUMAN, LOCATION, NUMERIC
- 50 finer classes
  - LOCATION: city, country, mountain...
  - HUMAN: group, individual, title...
  - ENTITY: animal, body, color, currency...

Learning Question Classifiers [Xin Li & Dan Roth, COLING 2002]
Question Answering [Dan Jurafsky, Stanford]
Factoid Answer based on Web Documents

- Part of the Answer Type Taxonomy

Learning Question Classifiers [Xin Li & Dan Roth, COLING 2002]
Question Answering [Dan Jurafsky, Stanford]
Factoid Answer based on Web Documents

• **Answer Type Detection**
  • Rules
    • Regular expression based rules
      • Who \{is|was|are|were\} PERSON
  • Question headword
    • Which **city** in China has the largest number of foreign financial companies?
    • What is the state **flower** of California?
  • Machine Learning
    • 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: Question words and phrases; Part-of-speech tags; Parse features (headwords); Named Entities; Related words

Question Answering [Dan Jurafsky, Stanford]
Factoid Answer based on Web Documents

- **Passage Retrieval**
  - Retrieve documents using query terms through search engines
  - Segment the documents into shorter units, like paragraphs.
  - Passage ranking, features
    - 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 passage
    - Longest sequence of question words
    - Rank of the document containing passage
    - ...

Question Answering [Dan Jurafsky, Stanford]
Factoid Answer based on Web Documents

• 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 CITY, tagger has to tag CITY
• Return the string with the right type:
  • How many bones in an adult human body? (Number)
  • The human skeleton is the internal framework of the body. It is composed of 270 bones at birth – this total decreases to 206 bones by adulthood after some bones have fused together.

Question Answering [Dan Jurafsky, Stanford]
Factoid Answer based on Web Documents

Knowledge Bases based QA

Web Documents based QA

Neil Armstrong
Apollo 11’s mission was to land two men on the moon. They also had to come back to Earth safely. Apollo 11 blasted off on July 16, 1969. Neil Armstrong, Edwin “Buzz” Aldrin and Michael Collins were the astronauts on Apollo 11. Jan 16, 2008

NASA - The First Person on the Moon
www.nasa.gov/audience/forstudents/k-4/stories/first-person-on-moon.html

Answer Sentence Selection
Answer Sentence Selection

- **Task**
  - Input:
    - a question
    - a set of candidate sentences
  - Output:
    - the correct sentence that contains the exact answer
    - can sufficiently support the answer choice
Answer Sentence Selection

- **Dataset**
  - QASent
    - Created using TREC-QA questions

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Dev</th>
<th>Test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># of ques.</td>
<td>94</td>
<td>65</td>
<td>68</td>
<td>227</td>
</tr>
<tr>
<td># of sent.</td>
<td>5,919</td>
<td>1,117</td>
<td>1,442</td>
<td>8,478</td>
</tr>
<tr>
<td># of ans.</td>
<td>475</td>
<td>205</td>
<td>248</td>
<td>928</td>
</tr>
<tr>
<td>Avg. len. of ques.</td>
<td>11.39</td>
<td>8.00</td>
<td>8.63</td>
<td>9.59</td>
</tr>
<tr>
<td>Avg. len. of sent.</td>
<td>30.39</td>
<td>24.90</td>
<td>25.61</td>
<td>28.85</td>
</tr>
</tbody>
</table>
# Answer Sentence Selection

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Reference</th>
<th>MAP</th>
<th>MRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Punyakanok (2004)</td>
<td>Wang et al. (2007)</td>
<td>0.419</td>
<td>0.494</td>
</tr>
<tr>
<td>Cui (2005)</td>
<td>Wang et al. (2007)</td>
<td>0.427</td>
<td>0.526</td>
</tr>
<tr>
<td>Wang (2007)</td>
<td>Wang et al. (2007)</td>
<td>0.603</td>
<td>0.685</td>
</tr>
<tr>
<td>H&amp;S (2010)</td>
<td>Heilman and Smith (2010)</td>
<td>0.609</td>
<td>0.692</td>
</tr>
<tr>
<td>Yao (2013)</td>
<td>Yao et al. (2013)</td>
<td>0.631</td>
<td>0.748</td>
</tr>
<tr>
<td>S&amp;M (2013)</td>
<td>Severyn and Moschitti (2013)</td>
<td>0.678</td>
<td>0.736</td>
</tr>
<tr>
<td>Shnarch (2013) - Backward</td>
<td>Shnarch (2013)</td>
<td>0.686</td>
<td>0.754</td>
</tr>
<tr>
<td>Yih (2013) - LCLR</td>
<td>Yih et al. (2013)</td>
<td>0.709</td>
<td>0.770</td>
</tr>
<tr>
<td>Yu (2014) - TRAIN-ALL bigram+count</td>
<td>Yu et al. (2014)</td>
<td>0.711</td>
<td>0.785</td>
</tr>
<tr>
<td>Feng (2015) - Architecture-II</td>
<td>Tan et al. (2015)</td>
<td>0.711</td>
<td>0.800</td>
</tr>
<tr>
<td>S&amp;M (2015)</td>
<td>Severyn and Moschitti (2015)</td>
<td>0.746</td>
<td>0.808</td>
</tr>
<tr>
<td>W&amp;I (2015)</td>
<td>Wang and Ittycheriah (2015)</td>
<td>0.746</td>
<td>0.820</td>
</tr>
<tr>
<td>Tan (2015) - QA-LSTM/CNN+attention</td>
<td>Tan et al. (2015)</td>
<td>0.728</td>
<td>0.832</td>
</tr>
<tr>
<td>dos Santos (2016) - Attentive Pooling CNN</td>
<td>dos Santos et al. (2016)</td>
<td>0.753</td>
<td>0.851</td>
</tr>
<tr>
<td>Wang et al. (2016) - Lexical Decomposition and Composition</td>
<td>Wang et al. (2016)</td>
<td>0.771</td>
<td>0.845</td>
</tr>
</tbody>
</table>

Bag of words, Word alignment, Dependency Tree Matching

Deep Neural Networks, LSTM

Answer Sentence Selection

- **Dataset**

<table>
<thead>
<tr>
<th></th>
<th>QASent</th>
<th>WikiQA</th>
</tr>
</thead>
<tbody>
<tr>
<td># of ques.</td>
<td>94 65 68</td>
<td>2,118 296 633</td>
</tr>
<tr>
<td># of sent.</td>
<td>5,919 1,117 1,442</td>
<td>20,360 2,733 6,165</td>
</tr>
<tr>
<td># of ans.</td>
<td>475 205 248</td>
<td>1,040 140 293</td>
</tr>
<tr>
<td>Avg. len. of ques.</td>
<td>11.39 8.00 8.63</td>
<td>7.16 7.23 7.26</td>
</tr>
<tr>
<td>Avg. len. of sent.</td>
<td>30.39 24.90 25.61</td>
<td>25.29 24.59 24.95</td>
</tr>
<tr>
<td># of ques. w/o ans.</td>
<td>1,245 170 390</td>
<td>1,245 170 390</td>
</tr>
</tbody>
</table>

Factoid Answer based on Web Documents

KB QnA

Who founded Apple?

Parsing and Transformation

Understanding

Knowledge Base

Steve Jobs - Founder

Apple Inc.

Inventor

Product

iPhone

iPad

Founder

Ronald Wayne

Founder

Steve Wozniak

Web QnA

Who first landed on the Moon?

Type Detection, NER Parsing and Candidate Ranking

Web Corpus

Apollo 11 was the spaceflight that landed the first humans on the Moon. Americans Neil Armstrong and Buzz Aldrin, on July 20, 1969, at 20:18 UTC.
Who first landed on the Moon?

Advantages:

• Generate better answer candidates
  • Entities in Freebase
  • Mentions of the same entity merged to one candidate

• Able to leverage entity information in Freebase
  • Semantic text relevance features for ranking
  • More fine-grained answer type checking

5% ~ 20% improvement in MRR

Open Domain Question and Answering via Semantic Enrichment [Huan Sun, et al., WWW 2015]
System Framework

Question
Who was the first American in space?

Sentence Collection
1. On May 5, 1961, Shepard piloted ...
2. Alan Shepard became the first American ...
3. ...

Candidate Generation
Via Entity Linking

Top-K Answers
1. Alan Shepard
2. Sally Ride
3. John Glenn
4. ...

Answer Candidate Pool
1. Freedom 7; 2. Alan Shepard
3. Sally Ride; 4. ...

Feature Generation & Ranking

Entity Info.

Open Domain Question and Answering via Semantic Enrichment [Huan Sun, et al., WWW 2015]
Experiments - Data

• **TREC Datasets** (well-formed questions)
  - Training: 1,700 (entity) questions (TREC 8-11)
  - Testing: 202 (entity) questions (TREC 12)

  **Example questions:**
  1. What are pennies made of?
  2. What is the tallest building in Japan?
  3. Who sang “Tennessee Waltz”? 

• **Bing Queries** (queries with question intent)
  - Training: 4,725 queries; Testing: 1,164 queries

  **Example queries:**
  1. the highest flying bird
  2. indiana jones named after
  3. designer of the golden gate bridge

Open Domain Question and Answering via Semantic Enrichment [Huan Sun, et al., WWW 2015]
Systems & Evaluation Metrics

• **QuASE** (Question Answering via Semantic Enrichment)
  • Includes other basic features (e.g., candidate freq.)
  • Ranker learner: MART (Multiple Additive Regression Trees)

• **Baselines**
  • AskMSR+ [Tsai+ ‘15] – Web-based QA system
  • SEMPRE [Berant+ ‘14] – Semantic parsing QA using Freebase

• **Evaluation Metrics**
  • MRR: Mean Reciprocal Rank
    • Determined by the top-ranked correct answer

Open Domain Question and Answering via Semantic Enrichment [Huan Sun, et al., WWW 2015]
Experiments – Results

MRR: Mean Reciprocal Rank

- QuASE: 0.65
- AskMSR+: 0.62
- SEMPRE: 0.14

Open Domain Question and Answering via Semantic Enrichment [Huan Sun, et al., WWW 2015]
Factoid Answer based on Tables

• What if answers cannot be found through KB and Web Documents?

Knowledge Bases  Tables  Web Documents

Structured  Semi-Structured  Unstructured
### Factoid Answer based on Tables

- What if answers cannot be found through KB and Web Documents?

**Q: Where is the largest brick dome?**

Below is a list of buildings that have held the title of the largest dome on their continent.

<table>
<thead>
<tr>
<th>Europe</th>
<th>Held record</th>
<th>Diameter</th>
<th>Name</th>
<th>Location</th>
<th>Builder</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1250 BC–1st century BC</td>
<td>14.5 m&lt;sup&gt;[1]&lt;/sup&gt;</td>
<td>Treasury of Atreus</td>
<td>Mycenae, Greece</td>
<td>City state of Mycenae</td>
<td>Corbel dome</td>
</tr>
<tr>
<td></td>
<td>1st century BC–19 BC</td>
<td>21.5 m&lt;sup&gt;[2]&lt;/sup&gt;</td>
<td>Temple of Mercury</td>
<td>Baiae, Italy</td>
<td>Roman Empire</td>
<td>First monumental dome&lt;sup&gt;[3]&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>1436–1881</td>
<td>45.52</td>
<td>Santa Maria del Fiore</td>
<td>Florence, Italy</td>
<td>Roman Catholic Archdiocese of Florence</td>
<td>Largest brick and mortar dome in the world till present. Octagonal dome.</td>
</tr>
</tbody>
</table>
Factoid Answer based on Tables

- What if answers cannot be found through KB and Web Documents?

<table>
<thead>
<tr>
<th>Web</th>
<th>Images</th>
<th>Videos</th>
<th>Maps</th>
<th>News</th>
<th>Explore</th>
</tr>
</thead>
<tbody>
<tr>
<td>13,100,000 RESULTS</td>
<td>Any time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phoenix Arizona Radio Stations</th>
<th>Frequency</th>
<th>Call Letters</th>
<th>City</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>550</td>
<td>KFYI</td>
<td>Phoenix</td>
<td>News/Talk</td>
</tr>
<tr>
<td></td>
<td>620</td>
<td>KTAR</td>
<td>Phoenix</td>
<td>Sports Talk</td>
</tr>
<tr>
<td></td>
<td>710</td>
<td>KBMG</td>
<td>Black Canyon City</td>
<td>Spanish Sports radio</td>
</tr>
<tr>
<td></td>
<td>740</td>
<td>KIDR</td>
<td>Phoenix</td>
<td>Spanish News/Talk</td>
</tr>
</tbody>
</table>

22 more rows, 1 more columns

Phoenix AM and FM Radio Station Guide
www.a2zphoenix.com/media/radio/
Factoid Answer based on Tables

- Knowledge Bases/Graphs
  - Structured but incomplete

- Unstructured Texts
  - Completely no structure

- Semi-Structured Tables
  - Rich: hundreds of millions tables [Lehmberg et al, WWW’16]

- Schema
  - Table caption
  - Column names
  - Table cells

<table>
<thead>
<tr>
<th>University</th>
<th>City</th>
<th>Province</th>
<th>Established</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Alberta</td>
<td>Calgary</td>
<td>Alberta</td>
<td>1906</td>
</tr>
<tr>
<td>University of Toronto</td>
<td>Toronto</td>
<td>Ontario</td>
<td>1827</td>
</tr>
<tr>
<td>University of Montreal</td>
<td>Montreal</td>
<td>Quebec</td>
<td>1878</td>
</tr>
</tbody>
</table>

List of universities in Canada
Factoid Answer based on Tables

Given:

Table Database

Question

What languages do people in France speak?

<table>
<thead>
<tr>
<th>Country</th>
<th>Capital</th>
<th>Location</th>
<th>Main Language</th>
<th>Currency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>Algiers</td>
<td>Africa</td>
<td>Arabic, French</td>
<td>Dinar</td>
</tr>
<tr>
<td>France</td>
<td>Paris</td>
<td>Europe</td>
<td>French</td>
<td>Euro</td>
</tr>
<tr>
<td>Hungary</td>
<td>Budapest</td>
<td>Europe</td>
<td>Hungarian</td>
<td>Forint</td>
</tr>
<tr>
<td>Singapore</td>
<td>Singapore</td>
<td>Asia</td>
<td>Malay, Chinese, Tamil</td>
<td>Singapore Dollar</td>
</tr>
</tbody>
</table>

The goal: to find a table cell containing answers.

Answer

French

Evidence

<table>
<thead>
<tr>
<th>Country</th>
<th>Main Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>France</td>
<td>French</td>
</tr>
</tbody>
</table>

Source: http://hasibul.info/gk/countries.php

Table Cell Search for Question Answering [Huan Sun, et al., WWW 2016]
Factoid Answer based on Tables

- Too many tables! How to find related ones?
  
  “What languages do people in France speak?”
  More than 100K tables contain “France”!
- How to precisely identify the answer cell?
  
  “What languages do people in France speak?”
  Capital? Main Language? Currency?

<table>
<thead>
<tr>
<th>Country</th>
<th>Capital</th>
<th>Currency</th>
<th>Main Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>Algiers</td>
<td>Dinar</td>
<td>Arabic</td>
</tr>
<tr>
<td>Egypt</td>
<td>Cairo</td>
<td>Pound</td>
<td>Arabic</td>
</tr>
<tr>
<td>France</td>
<td>Paris</td>
<td>Euro</td>
<td>French</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

A list of countries and their capital, language etc.

Table Cell Search for Question Answering [Huan Sun, et al., WWW 2016]
Factoid Answer based on Tables

• Question chain

Chain representation for “What languages do people in France speak?”: entity + question pattern

• Table cell chain

Graph representation of a table row:

Relational chain between “France” and “French”:

Table Cell Search for Question Answering [Huan Sun, et al., WWW 2016]
Factoid Answer based on Tables

**Question**
What languages do people in France speak?

**Candidate Chain Collection**
1. France -- > Table 1
2. France -- > Table 2
3. ...

**Top-K Chains**
1. Country -- MainLanguage
2. Mainly spoken in -- language
3. ...

**Pruned Chain Collection**
1. France -- > Table 2
2. France -- > Table k
3. ...

---

Table Cell Search for Question Answering [Huan Sun, et al., WWW 2016]
What languages do people in France speak?

**Step 1:**

**Candidate Chain Generation**

String match with table cells

<table>
<thead>
<tr>
<th>Country</th>
<th>Capital</th>
<th>Currency</th>
<th>Main Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>Algiers</td>
<td>Dinar</td>
<td>Arabic</td>
</tr>
<tr>
<td>Egypt</td>
<td>Cairo</td>
<td>Pound</td>
<td>Arabic</td>
</tr>
<tr>
<td>France</td>
<td>Paris</td>
<td>Euro</td>
<td>French</td>
</tr>
</tbody>
</table>

Generate an initial set of chains

{ France -> Table ID -> ? ;

France -> Table ID -> Capital -> ? , ... }

Table Cell Search for Question Answering [Huan Sun, et al., WWW 2016]
Factoid Answer based on Tables

Step 2: Coarse-grained Pruning via Snippets Matching

- Shallow features for each candidate chain
  1. Candidate chain side
     - Word vector using table title, caption, column names etc.
  2. Question side
     - Word vector using Bing snippets

- Select top-k candidate chains using shallow features

- Most irrelevant chains can be removed

Table Cell Search for Question Answering [Huan Sun, et al., WWW 2016]
Factoid Answer based on Tables

Step 3: Deep Chain Inference

Question: What languages do people in France speak?

Candidate chain:
1. France \(\rightarrow\) Table2 \(\rightarrow\) Table1
2. France \(\rightarrow\) Tablek
3. ...

Table Cell Search for Question Answering [Huan Sun, et al., WWW 2016]
Factoid Answer based on Tables

Step 3: Deep Chain Inference

Question

What languages do people in France speak?

Candidate chain

1. France \(\rightarrow\) Table2

2. France \(\rightarrow\) Tablek

3. …

Answer type: Column name w.r.t the answer cell
e.g., “Main Language”

Pseudo-predicate: Column name pair
e.g., \(<\text{country, main language}\>

Entity pairs: entity pairs in the two columns
e.g., \{ \(<\text{Spain, Spanish}>; <\text{Italy, Italian}>; … \) \}

Table Cell Search for Question Answering [Huan Sun, et al., WWW 2016]
Factoid Answer based on Tables

• Deep features
  • <question pattern, answer type>: DSSM-Type
  • <question pattern, pseudo-predicate>: DSSM-Predicate
  • <question pattern, entity pairs>: DSSM-EntityPairs

\[ R(Q, A) = \text{cosine}(y_Q, y_A) = \frac{y_Q^T y_A}{\|y_Q\| \|y_A\|} \]

e.g., DSSM-Type:

Semantic layer: \( y \)
Affine projection matrix: \( W_s \)
Max pooling layer: \( \nu \)
Max pooling operation

Convolutional layer: \( h_y \)
Convolution matrix: \( W_y \)
Word hashing layer: \( f_y \)
Word hashing matrix: \( W_f \)
Word sequence: \( x_q \)

Input: Question pattern

\( y_Q \)

Answer type

\( y_A \)
Question Sets

- **WebQuestions: WebQ**
  - Training: 3,778 (entity) questions
  - Testing: 2,032 (entity) questions

  **Example questions:**
  1. who did the voice for lola bunny?
  2. in what countries do people speak danish?

- **Bing Queries: BingQ**
  - Training: 4,725 queries
  - Testing: 1,164 queries

  **Example queries:**
  1. cherieff callie voice
  2. boeing charleston sc plant location
Table Sets

- WikiTables
  - Tables from Wikipedia and Wikipedia Infoboxes
  - ~5M Tables
Baselines and Metrics

- **TabCell**: Table Cell Search
  - Feature set: shallow features, deep features
  - Algorithm: MART (Multiple Additive Regression Trees)
- **Baselines**: Semantic parsing on Freebase
  - Sempre [Berant et al, EMNLP’13]
  - ParaSempre [Berant et al, ACL’14]
- **TabCell + ParaSempre**: simply combine their Top-1 results

**Evaluation Metrics**

- **Precision, Recall, F1**
  - # of answers in ground truth: N
  - # of true answers contained in top-1 table cell: M
  - Recall = M / N
  - Precision = 0 if M=0; 1 otherwise (b/c, only 1 table cell returned)
Factoid Answer based on Tables

- How Does TabCell Compare with ParaSempre?

![Bar chart showing F1 scores for TabCell, Sempre, and ParaSempre across WebQuestions and Bing Queries.](chart.png)

- For WebQuestions:
  - TabCell: 0.49
  - Sempre: 0.56
  - ParaSempre: 0.65

- For Bing Queries:
  - TabCell: 0.30
  - Sempre: 0.21
  - ParaSempre: 0.22

Difference:
- TabCell vs. Sempre: +0.06
- TabCell vs. ParaSempre: +0.16
- Sempre vs. ParaSempre: -0.01
Factoid Answer based on Tables

- Do Tables Complement KBs?
Factoid Answer based on Tables

• **Take-away Messages**
  • Tables contain rich knowledge to complement knowledge bases.
  • QA based on tables calls for deep understanding table semantics, e.g., column meaning and relations among columns.
Challenges in Web-based QA

4) An asteroid will hit on May 16, 2016 – followed by a black hole created by CERN. The end of the world is nigh, appaz (Picture Alamy) The world will be over by October 25, according to Pastor Ricardo Salazar, who's behind a series of very, very odd YouTube rants. Jan 4, 2016

5 reasons the world is going to end this year, probably on February 14...
metro.co.uk/.../5-reasons-the-world-is-going-to-end-this-year-probably-on-februa... Metro
Challenges in Web-based QA

Who will be the president in 2016

Who will win the 2016 U.S. presidential election?

<table>
<thead>
<tr>
<th></th>
<th>Latest</th>
<th>Buy Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hillary Clinton</td>
<td>65¢</td>
<td>1¢</td>
</tr>
<tr>
<td>Donald Trump</td>
<td>34¢</td>
<td>1¢</td>
</tr>
<tr>
<td>Joe Biden</td>
<td>4¢</td>
<td>NC</td>
</tr>
<tr>
<td>Bernie Sanders</td>
<td>3¢</td>
<td>2¢</td>
</tr>
</tbody>
</table>

PredictIt | Who will win the 2016 U.S. presidential election?
https://www.predictit.org/market/1234/who-will-win-the-2016-us-presidential-election
Challenges in Web-based QA

- **Question Understanding**
  - Rules are not always correct
  - “where is my refund”
  - location?
  - When and how to get refund
  - “when a cat loves a dog”
  - Date Time?
  - TV series
Question Answering for Testing Machine Intelligence
A Different Kind of Question Answering...

- Story comprehension (MCTest)
- Fill-in-the-blank questions (MSR sentence completion, DeepMind Q&A Dataset, Facebook Children Stories)
- Commonsense reasoning (Facebook bAbI)
- Quiz competition (Quiz Bowl, Jeopardy!)
- Standard test for measuring AI (AI2)
- Visual Question Answering
A Different Kind of Question Answering...

- Story comprehension (MCTest)
- Fill-in-the-blank questions (MSR sentence completion, DeepMind Q&A Dataset, Facebook Children Stories)
- Commonsense reasoning (Facebook bAbl)
- Quiz competition (Quiz Bowl, Jeopardy!)
- Standard test for measuring AI (AI2)
- Visual Question Answering
Story Comprehension – Early Work

  • Model the world knowledge
  • Understand natural language

  • A small reading comprehension dataset (3rd to 6th grade stories)
  • Find sentences to answer “who/what/when/where/why” questions
  • Simple BoW approach reaches 40% accuracy (~5% random)
MCTest: Reading Comprehension Test
[Richardson+, EMNLP-13]

- 660 children’s stories, 2,640 comprehension questions
- Data collection: Crowdsourcing via Amazon MTurk
  - No copyright issues, freely downloadable
- Fictional: Answers are found only in the story
- Grade-school level: limited vocabulary (8,000 words)
- Multiple-choice: objective/offline evaluation
- Open-Domain
Timmy liked to play games and play sports but more than anything he liked to collect things. He collected bottle caps. He collected sea shells. He collected baseball cards. He has collected baseball cards the longest. He likes to collect the thing that he has collected the longest the most. He once thought about collecting stamps but never did. His most expensive collection was not his favorite collection. Timmy spent the most money on his bottle cap collection.

1) Timmy liked to do which of these things the most?
   A) Collect things
   B) Collect stamps
   C) Play games
   D) Play sports

2) Which is Timmy's most expensive collection?
   A) Stamps
   B) Baseball Cards
   C) Bottle Cap
   D) Sea Shells

3) Which item did Timmy not collect?
   A) Bottle caps
   B) Baseball cards
   C) Stamps
   D) Sea shells

4) Which item did Timmy like to collect the most?
   A) Stamps
   B) Baseball cards
   C) Bottle caps
   D) Sea shells
Baselines

• Window Algorithm:
  • $S = \text{question} + \text{hypothesized answer}$
  • Score: best matching $|S|$-sized window in story
  • Answer with best score wins

• Distance Algorithm:
  • For each word in question, find distance in story to the nearest word in answer
  • Answer with lowest average distance wins

• MC500 Test Questions: W+D: 60.26% Accuracy
Fostered Research on a Variety of Approaches

- Lexical matching [Smith et al., 2015]
- Discourse processing [Narasimhan and Barzilay, 2015]
- Rules [Chen et al., 2015]
- Semantic frames [Wang et al., 2015]
- Memory Networks [Kapashi et al., 2015]
- Answer-entailing structures [Sachan et al., 2015]
- Attention-based CNNs [Yin et al., 2016]
- Parallel-Hierarchical NN [Trischler et al., 2016]
Answer-entailing structures [Sachan et al., 2015]

Fig. 1 of [Sachan et al., 2015]

Text: ... The restaurant had a special on catfish ... Alyssa enjoyed the restaurant’s special ...

Hypothesis: Alyssa ate Catfish at the restaurant.
(Question: What did Alyssa eat at the restaurant? Answer Candidate: Catfish)

- Latent structured SVMs with rich features
  - Lexical semantic features based on SENNA word vectors & WordNet
  - RST (Rhetorical Structure Theory) tags for cross-sentence relations
- Best accuracy: 67.83% (with multitask learning)
Parallel-Hierarchical NN [Trischler et al., 2016]

- Embed document and question/answer
- Combine multiple perspectives
  - Text semantic vectors
  - Sentential vectors
  - Sliding window based on words and dependency trees
- Accuracy: 71.0%

Fig. 1 of [Trischler et al., 2016]
Story Comprehension – Summary

- Simple baselines are strong (~60% vs. 25% random)
- ML-based “text matching” approaches are winning
  - LSSVMs + multitask leaning → 67.8%
  - Neural networks + word embedding → 70.0%

- Reasoning process is not easily interpretable
  - No explicit world knowledge or model has been used
  - Cannot provide explanations on why the answers are chosen
- Still room for improvement (the ceiling is 100%)
Fill-in-the-blank Quiz Questions

“At last she looked up with something _____ and defiant in her manner.”

a) reckless
b) solid
c) pallid
d) jovial
e) warm

Example from MSR Sentence Completion Challenge
Fill-in-the-blank Quiz Questions

• Motivation
  • Same high-level goal as MCTest
  • Seeking a more scalable way to collect data (e.g., vs. crowdsourcing)
    • MCTest dataset might be too small for supervised learning, especially for NN approaches

• High-level process
  • Pick a large corpus (e.g., news articles, stories)
  • Develop an (almost) automatic way to generate (fill-in-the-blank) questions
DeepMind Q&A Dataset [Hermann et al., NIPS-15]

- 93k CNN & 220k Daily Mail articles
- Bullet points (summary / paraphrases) → Cloze questions
  - Replacing one entity with a placeholder
  - ~4 questions per document
  - ~1M document / query / answer triples

- Datasets recreated by Kyunghyun Cho
  - http://cs.nyu.edu/~kcho/DMQA/
<table>
<thead>
<tr>
<th>Original Version</th>
<th>Anonymised Version</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Context</strong></td>
<td><strong>Anonymised Version</strong></td>
</tr>
<tr>
<td>The BBC producer allegedly struck by Jeremy Clarkson will not press charges</td>
<td>the <em>ent381</em> producer allegedly struck by <em>ent212</em> will not press charges against</td>
</tr>
<tr>
<td>against the “Top Gear” host, his lawyer said Friday. Clarkson, who hosted one</td>
<td>the “<em>ent153</em>” host, his lawyer said Friday. <em>ent212</em>, who hosted one of the</td>
</tr>
<tr>
<td>of the most-watched television shows in the world, was dropped by the BBC</td>
<td>most-watched television shows in the world, was dropped by the <em>ent381</em></td>
</tr>
<tr>
<td>Wednesday after an internal investigation by the British broadcaster found he</td>
<td>Wednesday after an internal investigation by the <em>ent180</em> broadcaster found he</td>
</tr>
<tr>
<td>had subjected producer Oisin Tymon “to an unprovoked physical and verbal attack.”</td>
<td>had subjected producer <em>ent193</em> “to an unprovoked physical and verbal attack.”</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Query</strong></td>
<td><strong>Anonymised Version</strong></td>
</tr>
<tr>
<td>Producer X will not press charges against Jeremy Clarkson, his lawyer says.</td>
<td>producer X will not press charges against <em>ent212</em>, his lawyer says.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Answer</strong></td>
<td><img src="" alt="Table" /></td>
</tr>
</tbody>
</table>
Word Counting Baselines

- Majority
  - Pick the most frequently observed entity in the $D$

- Exclusive majority
  - Same as Major, but the entity is not observed in $Q$
Symbolic Matching Models

- Frame-semantic parsing
  - Match PropBank triples \((x, V, y)\)
  - \(Q: \text{“X loves Sue”} \) vs. \(D: \text{“Kim loves Sue”}\)

- Word distance benchmark
  - Align the placeholder in \(Q\) with each possible entity in \(D\)
  - Sum the distances of each word in \(Q\) to nearest aligned words in \(D\)
Neural Network Models – Attentive Reader

\[ p(a|d, q) \propto \exp(W(a)g(d, q)) \]

- **s**: attention weight
- **y**: embedding

Mary went to England

[Hermann et al., NIPS-15. Fig 1a]
Neural Network Models – Impatient Reader

Mary went to England X visited England

[Hermann et al., NIPS-15. Fig 1b]
Accuracy

Simple window baseline is better than NL analysis

CNN
Max Freq: 33.2
Excl. Freq: 39.3
Frame-semantic: 40.2
Word Dist: 63
Att. Reader: 63.8
Impat. Reader: 50.9

Daily Mail
Max Freq: 25.5
Excl. Freq: 32.8
Frame-semantic: 35.5
Word Dist: 55.5
Att. Reader: 69
Impat. Reader: 68
A Thorough Examination... [Chen et al. ACL-16]

• Challenges & Questions
  • A clever way of creating large supervised data, but an artificial task
  • Unclear what level of reading comprehension needed

• Good News – The task is not really difficult!
  • An entity-centric classifier with simple features works fine
  • A variant of the Attentive Reader model achieves the new best result

• Bad News – The task is not really difficult!
  • Not much “comprehension” is needed
  • Probably have reached the ceiling
Accuracy

Frame-semantic  Word Dist.  Ent. Classifier  Att. Reader  Att. Reader+

CNN

Daily Mail

40.2  50.9  65.9  63.0  71.8

35.5  55.5  66.6  69.0  74.9
Analysis on 100 Examples from CNN

<table>
<thead>
<tr>
<th>Category</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact match</td>
<td>13%</td>
</tr>
<tr>
<td>Paraphrasing</td>
<td>41%</td>
</tr>
<tr>
<td>Partial clue</td>
<td>19%</td>
</tr>
<tr>
<td>Multiple sentences</td>
<td>2%</td>
</tr>
<tr>
<td>Coreference errors</td>
<td>8%</td>
</tr>
<tr>
<td>Ambiguous / hard (to human)</td>
<td>17%</td>
</tr>
</tbody>
</table>

- 25% questions are not answerable!
### Analysis on 100 Examples from CNN

<table>
<thead>
<tr>
<th>Category</th>
<th>Ratio</th>
<th>Classifier</th>
<th>NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact match</td>
<td>13%</td>
<td>13 (100.0%)</td>
<td>13 (100.0%)</td>
</tr>
<tr>
<td>Paraphrasing</td>
<td>41%</td>
<td>29 (70.7%)</td>
<td>39 (95.1%)</td>
</tr>
<tr>
<td>Partial clue</td>
<td>19%</td>
<td>14 (73.7%)</td>
<td>17 (89.5%)</td>
</tr>
<tr>
<td>Multiple sentences</td>
<td>2%</td>
<td>1 (50.0%)</td>
<td>1 (50.0%)</td>
</tr>
<tr>
<td>Coreference errors</td>
<td>8%</td>
<td>3 (37.5%)</td>
<td>3 (37.5%)</td>
</tr>
<tr>
<td>Ambiguous / hard (to human)</td>
<td>17%</td>
<td>2 (11.8%)</td>
<td>1 (5.9%)</td>
</tr>
</tbody>
</table>

- 25% questions are not answerable!
- NN handles paraphrases and lexical variations better.
Other Related Tasks & Datasets (1/3)

- MSR Sentence Completion Challenge [Zweig & Burges, 2011]
  - 1,040 sentences from five Sherlock Holmes novels
  - An infrequent word is chosen as the focus of the question
  - 4 alternates chosen by hand from 30 words suggested by LM
  - Random: 25%. Human: 91%. Current Best: 56% [Liu et al., ACL-15]

- Quiz Bowl: paragraph factoid questions [Iyyer et al., EMNLP-14]
  - Predict the entity described by the short paragraph
Other Related Tasks & Datasets (2/3)

- Facebook Children’s Book Test [Hill et al., ICLR-16]
  - 20 sentence as context
  - 21st sentence → Cloze question with 10 candidates

- ROCStories and Story Cloze Test Corpora
  [Mostafazadeh et al., NAACL-HLT-16]
  - 50k five-sentence commonsense stories
  - Given the first 4 sentences, select the correct ending
  - Designed to be 100% answerable by human judges
Other Related Tasks & Datasets (3/3)

- **Russian-language QA dataset** [Provided by Sergey Nikolenko]
- **Examples** (translated to English, courtesy of Sergey Nikolenko)
  - The professor later married a Ph.D. student Christina Maslach; she was the only person who explicitly objected. Which university was he a professor of?
  - An old Russian superstition recommends to pull weeds on the 18th of June. According to the second part of the same proverb, the 18th of June can also be considered favorable for THIS PROCESS. Name this process with a word of Latin origin.
  - A womanizer from a Viennese comic opera believes that IT reduces female resistance by a factor of four. Name IT.
Facebook bAbI Tasks [Weston et al., ICLR-16]

- 20 categories of simple commonsense reasoning tasks
  - A short description of agents moving around & passing objects
  - Followed by a simple question that can be answered based on the description
- 1,000/1,000 questions for training/testing

Task 3: Three Supporting Facts

John picked up the apple.
John went to the office.
John went to the kitchen.
John dropped the apple.
Where was the apple before the kitchen? A: office
Arguments for Creating bAbI Tasks

• Categorize different reasoning questions into skill sets

• Claims / Hopes:
  • Analyze model performance on different skills to study the strengths and weaknesses
  • Simple language and problems make the results easy to interpret
  • Each task checks a skill that a system should have
  • Mastering all the tasks is a prerequisite for any system with full text understanding and reasoning ability
Task Generation via a Simulated World

- States & properties of entities
- Actions an actor can take (e.g., go <loc>, get <obj>)
Memory Networks [Weston et al. 2014]

• Class of models instead of one model
• Key concepts
  • Explicit memory storage and index
  • Select memory for matching
• Basic components
  • Input feature map: sentence $x$ to an internal representation $I(x)$
  • Generalization: update memory $m$: $m_i = G(m_i, I(x), m)$, $\forall i$.
  • Output feature map: compute output $o$: $O(I(x), m)$
  • Response: decode $o$ to give a textual response $r = R(o)$
• Implementation could be very simple
Memory Networks for bAbI

- Input: embedding of simple bag of words
- Generalization: store embedding of sentences sequentially
- Output: find two supporting facts
  - 1\textsuperscript{st} supporting fact $s_1$ (max match score [dot product] with question $q$)
  - 2\textsuperscript{nd} supporting fact $s_2$ (max match score with $s_2$ & $q$)
- Response: rank possible answer words given the facts
  - Based on dot products of the word vector and the embedding of facts

75% accuracy; advanced variation achieves 93% accuracy
Task 19: Path Finding

The kitchen is north of the hallway.
The bathroom is west of the bedroom.
The den is east of the hallway.
The office is south of the bedroom.
How do you go from den to kitchen? *A: west, north*
How do you go from office to bathroom? *A: north, west*
Reasoning in Vector Space [Lee et al., ICRL-16]

- Decouple semantic parsing & logical reasoning

- Two vector-space reasoning models, inspired by Tensor Product Representation [Smolensky 1990/2006]
  - All entities are represented in $d$-dimensional unit vectors
  - Relation between two entities is described by matrix product (binding)
  - Inference (answering questions) is done by inner product

- 100% accuracy except Categories 5 & 16
  - Incorrect answers & ambiguity in facts
<table>
<thead>
<tr>
<th>#</th>
<th>Statements/Questions</th>
<th>Encodings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mary went to the kitchen.</td>
<td>$mk^T$</td>
</tr>
<tr>
<td>2</td>
<td>Mary got the football there.</td>
<td>$fm^T$</td>
</tr>
<tr>
<td>3</td>
<td>Mary travelled to the garden.</td>
<td>$mg^T$</td>
</tr>
<tr>
<td>4</td>
<td>Where is the football?</td>
<td></td>
</tr>
</tbody>
</table>

• Left-multiply by $f^T$ all statements prior to the current time 
  $(f^T \cdot mk^T, f^T \cdot fm^T, f^T \cdot mg^T)$

• Pick the most recent container where 2-norms are $\sim 1.0$ ($m^T$)

• If the container is an actor
  • Find the most recent container of the actor by left-multiplying by $m^T$ (Yields $g^T$)
  • Answer by the most recent container. ⇒ garden

• If the container is a location, return it as answer
Some Observations – Dataset Creation

• Synthetic or semi-synthetic
  ✓ Relatively easy to create large-scale datasets
  ✗ Datasets may have unexpected issues and thus more breakable

• Human generated or validated
  ✓ Datasets are more natural and real
  ✓ Could design specific reasoning tasks
  ✗ Less scalable, even with the help of crowdsourcing
Some Observations – Current Results

• Simple methods often provide strong baselines (vs. random)
• New methods give incremental improvement
• SOTA from statistical methods, but still far behind human

• Reasoning process is hard to interpret
  • For the ease of evaluation, being able to explain the decision process to human is not part of the metric
  • Not clear whether the solutions are general
Tutorial Summary – Part 1

- **Modern question answering applications**
  - Search engines evolve to handle question queries
  - Digital assistants address multi-turn QA
  - Business analytics service adopt natural language QA interface

- **Pioneer work on question answering machines**
  - Similar problems & applications
  - Limited success, often ad-hoc solutions
  - Constrained by data size, computational power & models
Tutorial Summary – Part 2

• Open-domain factoid question answering with KB
  • Large-scale knowledge bases as the sole information source
  • Find entities or properties of entities in KB to answer questions

• Mainstream approach – semantic parsing of questions
  • Map natural language questions to logical forms / structured queries
  • Accurate answers when parse & KB is complete and correct
  • Able to explain how the answers are derived
  • Challenges: language mismatch, large search space, compositionality
Tutorial Summary – Part 3

• Open-domain factoid question answering with the Web
  • Leverage Web redundancy – commonly asked facts stated frequently in various Web documents
  • Recent approaches to incorporate structured (KB) and semi-structured (Web tables) information sources

• Challenges
  • Difficult in handling domain-specific or tail questions
  • Deeper understanding of questions
Tutorial Summary – Part 4

- **Question answering for testing machine intelligence**
  - Designed to test AI; Not to fulfill users’ information need
  - A long-standing research strategy

- **Introduced recently proposed tasks**
  - Story comprehension (multiple-choice questions)
  - Fill-in-the-blank questions (find entities)
  - Commonsense reasoning (find answer words)

- **Challenges**
  - Having a well-designed and large dataset/task
Future

- Conversational intelligence supported by QA
  - No longer an independent task
  - Integrated naturally in a conversational system

- Multi-modal interaction
  - Visual question answering
  - Virtual tour guide
References (1/4)

References (2/4)

References (3/4)

- Richardson & White. “Supporting synchronous social Q&A throughout the question lifecycle.” WWW-2011.
- Smith et al. “A strong lexical matching method for the machine comprehension test.” EMNLP-2015
References (4/4)

- Yin et al. “Attention-Based Convolutional Neural Network for Machine Comprehension.” NAACL-2016 Workshop: Human-Computer QA.