Question Answering over Knowledge Bases:
Entity, Text, and System Perspectives

Wanyun Cui
Fudan University
Backgrounds

- Question Answering (QA) systems answer questions posed by humans in a natural language.
Backgrounds

• Question Answering (QA) systems answer questions posed by humans in a natural language.

IBM Watson

Watson prevailed over the human competitors and received the first place prize of $1 million.
Backgrounds

• Question Answering (QA) systems answer questions posed by humans in a **natural language**.

Apple Siri

Siri is the star of WWDC 2016: one, it’s supposed to launch on the Mac. And two, it’s about to get a lot more powerful.
Backgrounds

Google
Google Now

Apple
Siri

Amazon
Alexa

Microsoft
Cortana
Backgrounds: Applications

Health care

Bumrungrad Hospital and Watson Improve Cancer Care.

Digital assistant

Craig Federighi speaks during the Apple WWDC 2016.

Natural language search

Cortana supports natural language search for files on your computer.
Backgrounds: Resources for QA

Web documents  QA community  Search engine  Encyclopedia  Knowledge base

Plain text  Structured data (RDF)
Backgrounds

• Knowledge bases provide **accurate and concise** results.
  • A computer can easily access the structured data.

• The emerging of billion scale knowledge bases increase the **coverage** of questions knowledge bases can answer.
Backgrounds

- Knowledge bases provide **accurate and concise** results.
  - A computer can easily access the structured data.

- The emerging of billion scale knowledge bases increase the **coverage** of questions knowledge bases can answer.

- Open-domain QA over knowledge bases is feasible
Architecture

- QA is a complicated project, which involves systematic and multi-level semantic understanding.
QA is a complicated project, which involves systematic and multi-level semantic understanding.

**Underlying Technologies:**
- **Semantics of Entities**
- **Semantic Community Search**
- **Co-occurrence Network**

**Semantics of Short Text/Sentence**
- **Verb Pattern**
- **Context-Aware Conceptualization**
- **Semantic-Aware Word Embedding**

**Core Application:** Question Answering System
- **Question Templates**
- **Deep Learning**

**Entities:**
- Provides basic computing units of entities

**Short text/sentences:**
- Studies syntactic features using knowledge bases
- Connects low level techniques to high level applications

**QA system:**
- Core application
Entities
Semantic community search

• In a semantic network, the community of a word represents its semantics.

  - Community of “pot” in WordNet
    • Pot isA vessel and container
    • Pot is containerful (adj)
    • Pot, bowl, dish belong to the same category
Semantic community search

• In a semantic network, the community of a word represent its semantics.

• Goal: Find the semantic community of a given word

• Applications:
  • Semantic relatedness
  • Word sense disambiguation
  • Semantic expansion
Application: Semantic relatedness

- Pot & bowl
  - Same community
  - High relatedness
Application: Semantic relatedness

- Pot & bowl
  - same community
  - High relatedness

- Pot & plate
  - Overlap: 2 nodes
  - Medium relatedness
Application: Semantic relatedness

- Pot & bowl
  - same community
  - High relatedness
- Pot & plate
  - Overlap: 2 nodes
  - Medium relatedness
- Pot & tube
  - Overlap: 1 node
  - Low relatedness
Semantic community search

• Local search: a linear algorithm which iteratively adds new nodes from the local graph.
  • Estimation of nodes
  • Estimation of edges

\[ |V_{\geq k}| \sum_{t=1}^{\omega} t \times q_t = n \sum_{i=k}^{\omega} p_i \sum_{t=1}^{\omega} t \times q_t \]

• Performance increases of about 100-1000 times.

Wanyun Cui, et al., Local Search of Communities in Large Graphs, SIGMOD 2014
Towards deeper study

One word

Multiple senses

Multiple communities
Towards deeper study

Further research is needed.
Sense disambiguation

• Find the overlapping communities of a given entity.
• Performance increases over 1000 times.

4 overlapping communities for “vessel”
• C1: small boat
• C2: bucket
• C3: big boat
• C4: tablewares

Wanyun Cui, et al., Online Search of Overlapping Communities, SIGMOD 2013
Short Text/Sentence

Core Application: Question Answering System
- Question Templates
- Deep Learning

Semantics of Short Text/Sentence
- Verb Pattern
- Context-Aware Conceptualization
- Semantic-Aware Word Embedding

Underlying Technologies: Semantics of Entities
- Semantic Community Search
- Sense Disambiguation
- Co-occurrence Network

Deep Learning
- Semantic-Aware Word Embedding
- Sense Disambiguation
- Co-occurrence Network
- Core Application: Question Answering System
• “Go watch Rick and Morty if you haven’t already.”
  - TV show ✓ Movie ✓ Restaurant ✗ Instrument ✗

• Chinese eat Tang-Yuan in the cold December.
  - Food ✓ Meal ✓ Person ✗ Location ✗
• Verb pattern
  • To represent semantics of verbs by objects’ concept
  • Conceptualized: verb $c$concept
    • E.g. Watch $c$movie, Eat $c$food
  • Idiom patterns: verb $i$object
    • E.g. Kick $i$ass, Beat $i$retreat

• Principles
  • Generality: a verb pattern covers one meaning of the verb.
  • Specificity: a verb pattern matches its corresponding verb phrases.
Verb pattern

• Approach: Minimum Description Length (MDL)

\[
\arg\min_f \sum_P P(p) L(p)
\]

• Verb pattern is a compressed representation of verb phrases.
• Pattern assignment provides a compact description of phrases
• Pattern assignment captures the underlying verb semantics well.

Wanyun Cui, et al., Verb Pattern: A Probabilistic Semantic Representation on Verbs, (AAAI 2016)
Question Answering System

Underlying Technologies: Semantics of Entities
Semantic Community Search
Co-occurrence Network
Semantics of Short Text/Sentence
Verb Pattern
Context-Aware Conceptualization
Semantic-Aware Word Embedding
Core Application: Question Answering System
Question Templates
Deep Learning
Semantic-Aware Word Embedding
Sense Disambiguation
Co-occurrence Network

Knowledge Base
The question is understood by knowledge bases, including entity linking and predicate inference.

KB-based QA systems return answers by lookup entity and predicate in the knowledge base.
Template representation: Replace the entity in question by its concepts

A template represents complete intent of the question.

Key problem: collect templates and identify their corresponding predicates
Predicate Inference

- Automatic learning of templates and their corresponding predicates from QA corpora.
  - Generative model on probabilistic graph.
  - $P(e|q)$ entity recognition
  - $P(t|e,q)$ template extraction
  - $P(p|t)$ predicate inference
  - $P(v|e,p)$ value lookup
Question Templates

- 27126355 question templates
- 2782 question intents

System interface

<table>
<thead>
<tr>
<th>Category</th>
<th>Date</th>
<th>Entity</th>
<th>Location</th>
<th>Human</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ratio</td>
<td>39.4%</td>
<td>59.0%</td>
<td>13.8%</td>
<td>22.7%</td>
<td>0.6%</td>
</tr>
</tbody>
</table>

Wanyun Cui, et al., KBQA: An Online Template Based Question Answering System over Freebase, *(IJCAI 2016)*, demo

Results on QALD-5
KB-based QA + deep learning

- Deep learning has advantages in learning sequential data
  - CNN

Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Base
KB-based QA + deep learning

• Deep learning: entity type, answer path, answer context
Current issues of KB-based QA

• Lacks of high quality training data
  • WebQuestions: 3778 qa pairs
  • QALD: 100 qa pairs

• Missing facts in knowledge bases
  • Knowledge base completion
  • Reasoning

• Combination of KB-based QA and IR-based QA
  • KB-based: limited relations, precise values
  • IR-based: unlimited relations, fuzzy values
Thank you