Question Answering over Freebase with Multi-Column Convolutional Neural Networks

Li Dong¹, Furu Wei², Ming Zhou², Ke Xu¹
¹SKLSDE, Beihang University, Beijing, China
²Microsoft Research, Beijing, China
Question Answering over Freebase

- **Freebase**
  - Large-scale knowledge base
  - A rich resource to answer open-domain questions

  **Question:**
  when did Avatar release in UK

  ![Freebase](image)

  **Answer:** 2009-12-17

- **Challenge**
  - Natural language questions ~ structured semantics of Freebase
  - How to bridge the gap?
Mainstream Methods (1/2)

- **Semantic parsing** (Berant et al., 2013; Bao et al., 2014; etc.)
  - Question $\rightarrow$ **Formal Meaning Representation** $\rightarrow$ Structured Queries $\rightarrow$ **Freebase** $\rightarrow$ Answer
  
- **Example**
  - Utterance: Which college did Obama go to
  - Logical form: (and (Type University) (Education BarackObama))
  - Denotation: Occidental College, Columbia University

- **Challenges**
  - Huge search space
  - Lexical triggers

Example is borrowed from the website of SEMPRE
Mainstream Methods (2/2)

- Information extraction over knowledge base
  - 1. Retrieve candidate answers from Freebase
  - 2. Extract features
  - 3. Classification / Ranking

---

**Question**  
**Candidate Answers**  
**Features**  
**Classifier**  
**Correct Answer**  

(Yao and Van Durme, 2014)

---

**Question**  
**Candidate Embedding**  
**Candidate Answers**  
**Sum of Word Embeddings**  
**Correct Answer**  

(Bordes et al., 2014a; 2014b)
Proposed Method

- Question answering -> Constraint matching
  - Answer type, answer path (relation), answer context
- Question understanding with convolutional neural networks
when did Avatar release in UK
Embedding Candidate Answers

- Learn vector representations for candidate answers
  - (Bordes et al., 2014a; Bordes et al., 2014b)
- Answer path
  - relations between the candidate node and the entity asked in question
  - $\text{avg}(r_1, r_2, ..., r_n)$: average of relation embeddings
- Answer context
- Answer type
Embedding Candidate Answers

- Learn vector representations for candidate answers
  - (Bordes et al., 2014a; Bordes et al., 2014b)

- Answer context
  - 1-hop entities and relations connected to the answer path
  - $\text{avg}(c_1, c_2, \ldots, c_n)$: average of context entity and relation embeddings

- Answer path

- Answer type
Embedding Candidate Answers

- Learn vector representations for candidate answers
  - (Bordes et al., 2014a; Bordes et al., 2014b)

- Answer type
  - common.topic.notable_types, value type (e.g., float, string, datetime)
  - \( \text{avg}(t_1, t_2, ..., t_n) \): average of type embeddings

- Answer path
- Answer context
when did Avatar release in UK
when did Avatar release in UK

Avatar release date in UK:
2009-12-17
Model Training

- Negative instance \( a' \) is randomly sampled from the set of candidate answers
- Hinge loss for \((q, a)\) and \((q, a')\)
  \[
  l(q, a, a') = (m - S(q, a) + S(q, a'))_+ , \quad \text{where } (z)_+ = \max\{0, z\}
  \]
- Objective function
  - \( A_q \): set of correct answers
  - \( R_q \subseteq C_q \setminus A_q \): set of wrong answers
  \[
  \min \sum_q \frac{1}{|A_q|} \sum_{a \in A_q} \sum_{a' \in R_q} l(q, a, a')
  \]
- Back-propagation, AdaGrad, max-norm regularization
Inference (During Test)

1. Link to entity in Freebase
2. Retrieve candidates
   - Avatar
3. Compute vector representations
4. Compute scores

Candidate Answers
(2-hop entities/attributes)

when did Avatar release in UK

Ranker → Answer
Matching Score

Multi-column
Convolutional Neural Networks
Inference (During Test)

- If there are more than one correct answers
  - Use the margin $m$ in objective function as threshold
  - Candidates whose scores are not far from the best answer are regarded as predicted results

\[
\hat{A}_q = \{ \hat{a} \mid \hat{a} \in C_q \ \text{and} \ \max_{a' \in C_q} \{ S(q, a') \} - S(q, \hat{a}) < m \}
\]
Question Paraphrases for Multi-Task Learning

- Question understanding results of paraphrases should be same
  - who is the father of A
  - who is A’s father

- So, the vectors of paraphrases computed by neural networks should be similar
  - Hinge loss
  - Negative instance is randomly sampled

\[ l_p (q_1, q_2, q_3) = \sum_{i=1}^{3} \left( m_p - f_i (q_1)^T f_i (q_2) + f_i (q_1)^T f_i (q_3) \right)_+ \]
Experiments

- **WebQuestions** (Berant et al., 2013)
  - wh- questions collected by querying Google Suggest API
  - Annotated in Amazon Mechanical Turk
  - Train: 3023, Dev: 755, Test: 2032

  **Example**
  - Question: what is the name of justin bieber brother?
  - Answers: {Jazmyn Bieber, Jaxon Bieber}

- **Paraphrases** (Fader et al., 2013)
  - Collected from the WikiAnswers website
  - ~2.4M questions, grouped into ~355k paraphrase clusters
Experimental Results

- Better or comparable results than baseline methods

<table>
<thead>
<tr>
<th>Method</th>
<th>F1</th>
<th>P@1</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Berant et al., 2013)</td>
<td>31.4</td>
<td>-</td>
</tr>
<tr>
<td>(Berant and Liang, 2014)</td>
<td>39.9</td>
<td>-</td>
</tr>
<tr>
<td>(Bao et al., 2014)</td>
<td>37.5</td>
<td>-</td>
</tr>
<tr>
<td>(Yao and Van Durme, 2014)</td>
<td>33.0</td>
<td>-</td>
</tr>
<tr>
<td>(Bordes et al., 2014a)</td>
<td>39.2</td>
<td>40.4</td>
</tr>
<tr>
<td>(Bordes et al., 2014b)</td>
<td>29.7</td>
<td>31.3</td>
</tr>
<tr>
<td>MCCNN (our)</td>
<td>40.8</td>
<td>45.1</td>
</tr>
</tbody>
</table>
Model Analysis

- Ablation experiments
  - w/o path/type/context
    - without the specific column
  - w/o multi-column
    - tying parameters of multiple columns
  - w/o paraphrase
    - without question paraphrases
  - 1-hop
    - 1-hop paths to generate candidates

<table>
<thead>
<tr>
<th>Setting</th>
<th>F1</th>
<th>P@1</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>40.8</td>
<td>45.1</td>
</tr>
<tr>
<td>w/o path</td>
<td>32.5</td>
<td>37.1</td>
</tr>
<tr>
<td>w/o type</td>
<td>37.7</td>
<td>40.9</td>
</tr>
<tr>
<td>w/o context</td>
<td>39.1</td>
<td>41.0</td>
</tr>
<tr>
<td>w/o multi-column</td>
<td>38.4</td>
<td>41.8</td>
</tr>
<tr>
<td>w/o paraphrase</td>
<td>40.0</td>
<td>43.9</td>
</tr>
<tr>
<td>1-hop</td>
<td>29.3</td>
<td>32.2</td>
</tr>
</tbody>
</table>
Salient Question Words Detection

- **Saliency score**
  - How much a word affects question understanding
  - Replace a word with stop words, how much the vectors are affected

```
where is the microsoft located
<stopword> is the microsoft located
```

```
vector distance is salience score
```
Observations
- wh- words
- nouns dependent of the wh- words
  - type/country/leader
- verbs
  - speak/located
Future Work

- Question answering over unstructured text

Titanic is a 1997 American epic romantic disaster film directed, written, co-produced, and co-edited by James Cameron. ...

Who is the director of Titanic?

Type: Person
THANKS!
<table>
<thead>
<tr>
<th>Column 1 (Answer Path)</th>
<th>Column 2 (Answer Type)</th>
<th>Column 3 (Answer Context)</th>
</tr>
</thead>
<tbody>
<tr>
<td>what to do in hollywood can this weekend</td>
<td>where be george washington originally from</td>
<td>where do charlie draw go to college</td>
</tr>
<tr>
<td>what to do in midland tx this weekend</td>
<td>where be george washington carver from</td>
<td>where do kevin love go to college</td>
</tr>
<tr>
<td>what to do in cancun with family</td>
<td>where be george bush from</td>
<td>where do pauley perrette go to college</td>
</tr>
<tr>
<td>what to do at fairfield can</td>
<td>where be the thame river source</td>
<td>where do kevin jame go to college</td>
</tr>
<tr>
<td>what to see in downtown asheville nc</td>
<td>where be the main headquarters of google</td>
<td>where do charlie draw go to high school</td>
</tr>
<tr>
<td>what to see in toronto top 10</td>
<td>in what town do ned kelly and he family grow up</td>
<td>where do draw bree go to college wikianswer</td>
</tr>
<tr>
<td>who found collegehumor</td>
<td>who be the leader of north korea today</td>
<td>who be judy garland father</td>
</tr>
<tr>
<td>who found the roanoke settlement</td>
<td>who be the leader of syrium now</td>
<td>who be clint eastwood date</td>
</tr>
<tr>
<td>who own skywest</td>
<td>who be the leader of cuba 2012</td>
<td>who be emma stone father</td>
</tr>
<tr>
<td>who start mary kay</td>
<td>who be the leader of france 2012</td>
<td>who be robin robert father</td>
</tr>
<tr>
<td>who be the owner of kfc</td>
<td>who be current leader of cuba today</td>
<td>who miley cyrus engage to</td>
</tr>
<tr>
<td>who own wikimedia foundation</td>
<td>who be the minority leader of the house of representative now</td>
<td>who be chri cooley marry to</td>
</tr>
<tr>
<td>what type of money do japanese use</td>
<td>what be the official language of paraguay</td>
<td>what be the timezone in vancouver</td>
</tr>
<tr>
<td>what kind of money do japanese use</td>
<td>what be the local language of israel</td>
<td>what be my timezone in california</td>
</tr>
<tr>
<td>what type of money do jamaica use</td>
<td>what be the four official language of nigerium</td>
<td>what be los angeles california time zone</td>
</tr>
<tr>
<td>what type of currency do brazil use</td>
<td>what be the official language of jamaica</td>
<td>what be my timezone in oklahoma</td>
</tr>
<tr>
<td>what type of money do you use in cuba</td>
<td>what be the dominant language of jamaica</td>
<td>what be my timezone in louisiana</td>
</tr>
<tr>
<td>what money do japanese use</td>
<td>what be the official language of brazil now</td>
<td>what be the time zone in france</td>
</tr>
</tbody>
</table>