Learning to Transform, Combine, and Reason in Open-Domain Question Answering

Mostafa Dehghani, Hosein Azarbonyad, Jaap Kamps, and Maarten de Rijke

University of Amsterdam
Open Domain Question Answering

- Users seek direct answers to complex questions from large open-domain knowledge sources like the Web.
- Most open-domain question answering systems:
  - Use a search engine to retrieve a set of candidate documents.
  - Select one or a few of them as the context.
  - Apply reading comprehension models to extract answers.
Example

Q. Who is the Spanish artist, sculptor and draughtsman famous for co-founding the Cubist movement?

Answer: Pablo Picasso
Example

Q. Who is the Spanish artist, sculptor and draughtsman famous for co-founded the Cubist movement?

Answer: Pablo Picasso

D1: [...] Georges Braque, a major 20th-century artist, sculptor and collagist who co-founded the Cubist movement, was mobilized into the army in 1914 and [...]

Example

Q. Who is the Spanish artist, sculptor and draughtsman famous for co-founding the Cubist movement?

Answer: Pablo Picasso

D1. [...] Georges Braque, a major 20th-century artist, sculptor and collagist who co-founded the Cubist movement, was mobilized into the army in 1914 and [...]

D3. [...] and through this artistic collaboration, Pablo Picasso and Georges Braque founded Cubism, a style of painting that shattered traditional forms of artistic representation [...]

D12. [...] Picasso is a notable painter, sculptor, printmaker, and decorative artist who was born in the city of Málaga [...]

D66. [...] When the Spanish city of Malaga wanted to clean up its sleazy portside district, it called in D*Face, the big street artist [...]

Q. Who is the Spanish artist, sculptor and draughtsman famous for co-founding the Cubist movement?

Answer: Pablo Picasso

D1. [...] Georges Braque, a major 20th-century artist, sculptor and collagist who co-founded the Cubist movement, was mobilized into the army in 1914 and [...]

D3. [...] and through this artistic collaboration, Pablo Picasso and Georges Braque founded Cubism, a style of painting that shattered traditional forms of artistic representation [...]

D12. [...] Picasso is a notable painter, sculptor, printmaker, and decorative artist who was born in the city of Málaga [...]

D66. [...] When the Spanish city of Malaga wanted to clean up its sleazy portside district, it called in D*Face, the big street artist [...]

Picasso was born in Spain
Example

Q. Who is the Spanish artist, sculptor and draughtsman famous for co-founding the Cubist movement?

Answer: Pablo Picasso

D1. [...] Georges Braque, a major 20th-century artist, sculptor and collagist who co-founded the Cubist movement, was mobilized into the army in 1914 and [...]

D3. [...] and through this artistic collaboration, Pablo Picasso and Georges Braque founded Cubism, a style of painting that shattered traditional forms of artistic representation [...]

D12. [...] Picasso is a notable painter, sculptor, printmaker, and decorative artist who was born in the city of Málaga [...]

D66. [...] When the Spanish city of Malaga wanted to clean up its sleazy portside district, it called in D*Face, the big street artist [...]

Picasso, who was a Spanish artist co-founded Cubism

Picasso was born in Spain
Some questions, however, require taking a broader context into account, e.g., by considering low-ranked documents that are not immediately relevant, combining information from multiple documents, and reasoning over multiple facts from these documents to infer the answer.
Need for considering a **larger** set of documents?

- This can potentially come at the cost of:
  - **Efficiency**, as we need to process a larger input.
  - **Effectiveness**, as there will be more noisy and non-relevant documents when we go down the ranked list of candidate documents.
TraCRNet!

- **TraCRNet**: *Transform, Combine, and Reason*
  - Pronounced “Tracker Net”!
  - Is based on *Transformer* architecture.
  - *Efficiently* operates over a larger set of candidate documents.
  - Combining the evidence from all documents during *multiple steps of reasoning*.
  - *Robust against noise* from low-ranked non-relevant documents.
Recurrent Models
Transformer (general idea!)

Transformer (general idea!)

Universal Transformer (general idea!)

TraCRNet Architecture

- **Input Encoding**
  - Transformer Encoder
  - Transformer Encoder
  - Transformer Encoder

- **Multi-hop Reasoning**
  - Universal Transformer Encoder

- **Output Decoding**
  - Transformer Decoder
  - Answer

- **Encoding**
  - **Document level encoding**
  - **Token level encoding**
TraCRNet Architecture
On the Architectural Choices

- Transformer-based (No recurrence!)
  - Easily parallelizable $\rightarrow$ FAST!

- Transformer Encoder for Input Encoding
  - Global receptive field for dealing with long documents (token level encoding)
  - **Memorizing** the meaning of infrequent words
On the Architectural Choices

● Universal Transformer for Multi-hop Reasoning
  ○ Recurrence in depth $\rightarrow$ a strong inductive bias that enables the model to learn iterative or recursive transformations $\rightarrow$ chain of reasoning!

● Transformer + Universal Transformer
  ○ Encoding local information + Combining global information
  ○ Enough memorization + Ability of generalization.
Experiments

Datasets

- Two datasets
  - SearchQA
    - 140k question-answer pairs crawled from J!Archive, and augmented with text snippets retrieved using the Google search engine.
  - Quasar-T
    - 43k open-domain trivia questions and their answers obtained from various internet sources. The set of candidate documents for each question is retrieved using Lucene from the ClueWeb09 corpus as the background corpus.
Experiments

Main Results

- TraCRNet has the chance of using “full” use of the information of “all” the candidate documents.

- TraCRNet generates the answer instead of extracting it.

<table>
<thead>
<tr>
<th>model</th>
<th>SearchQA</th>
<th>Quasar-T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EM</td>
<td>F1</td>
</tr>
<tr>
<td>BiDAF [29]</td>
<td>28.6</td>
<td>34.6</td>
</tr>
<tr>
<td>R³ [34]</td>
<td>49.0</td>
<td>55.3</td>
</tr>
<tr>
<td>Wang et al. [35]</td>
<td>57.0</td>
<td>63.2</td>
</tr>
<tr>
<td>Lin et al. [26]</td>
<td><strong>58.8</strong></td>
<td>64.5</td>
</tr>
<tr>
<td>TraCRNet</td>
<td>52.9</td>
<td><strong>65.1</strong></td>
</tr>
<tr>
<td>Human Performance</td>
<td>43.9</td>
<td>–</td>
</tr>
</tbody>
</table>
Experiments
Effect of multihop reasoning

$\text{TraCRNet}_{\text{no-mhr}}^d$

○ The decoder has access to document-level representations from the encoder.

$\text{TraCRNet}_{\text{no-mhr}}^w$

○ The pooling operation is removed and the decoder has access to word-level representations from the encoder.

Table 3: Performance of TraCRNet with and without the Multihop Reasoning layer; numbers in parenthesis indicate percentage of performance loss

<table>
<thead>
<tr>
<th>model</th>
<th>SearchQA</th>
<th></th>
<th>Quasar-T</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>EM</td>
<td>F1</td>
<td>EM</td>
</tr>
<tr>
<td>TraCRNet</td>
<td>52.9</td>
<td>65.1</td>
<td></td>
<td>43.2</td>
</tr>
<tr>
<td>TraCRNet$_{\text{no-mhr}}^d$</td>
<td>48.6 (−8%)</td>
<td>61.7 (−5%)</td>
<td></td>
<td>36.4 (−16%)</td>
</tr>
<tr>
<td>TraCRNet$_{\text{no-mhr}}^w$</td>
<td>50.2 (−5%)</td>
<td>59.3 (−9%)</td>
<td></td>
<td>38.1 (−12%)</td>
</tr>
</tbody>
</table>
Back to our Example

Q. Who is the Spanish artist, sculptor and draughtsman famous for co-founding the Cubist movement?

Answer: Pablo Picasso

D1. [...] **Georges Braque**, a major 20th-century artist, sculptor and collagist who co-founded the Cubist movement, was mobilized into the army in 1914 and [...]

D3. [...] and through this artistic collaboration, **Pablo Picasso** and **Georges Braque founded Cubism**, a style of painting that shattered traditional forms of artistic representation [...]

D12. [...] **Picasso** is a notable painter, sculptor, printmaker, and decorative artist who was born in the city of Málaga [...]

D66. [...] When the Spanish city of Malaga wanted to clean up its sleazy portside district, it called in D*Face, the big street artist [...]

Picasso, who was a Spanish artist co-founded Cubism

Picasso was born in Spain
Multihop Reasoning

(a) Attention distribution when transforming the document at rank 12, in step#3 of multihop reasoning.

(b) Attention distribution when transforming the question, in step#7 of multihop reasoning.

Figure 3: Visualization of multi-head self-attention on Multihop Reasoning layer of TraCRNet. (Best viewed in color.)
Multihop Reasoning

(a) Attention distribution when transforming the document at rank 12, in step #3 of multihop reasoning.

(b) Attention distribution when transforming the question, in step #7 of multihop reasoning.

Figure 3: Visualization of multi-head self-attention on Multihop Reasoning layer of TraCRNet. (Best viewed in color.)
Multihop Reasoning

(a) Attention distribution when transforming the document at rank 12, in step#3 of multihop reasoning.

(b) Attention distribution when transforming the question, in step#7 of multihop reasoning.

Figure 3: Visualization of multi-head self-attention on Multihop Reasoning layer of TraCRNet. (Best viewed in color.)
Experiments
Impact of number of documents

- Train and evaluate models using different numbers of candidate documents associated with each question.

- This shows that, for some questions, the correct answer can only be extracted when considering information from low-ranked documents that are not immediately relevant to the question.

Figure 4: Performance in terms of F1 of TraCRNet and baselines ($R^3$ [34] and Lin et al. [26]'s model) with different numbers of candidate documents on Quasar-T dataset.
THANK YOU