Trained Trigger Language Model for Sentence Retrieval in Question Answering Systems

Saeedeh Momtazi
Spoken Language Systems
Saarland University

saeedeh.momtazi@lsv.uni-saarland.de

Research Advisor: Dietrich Klakow

PIRE
Outline

- Question Answering System
- Language Models for Information Retrieval
- Trained Triggering Model
- Results
- Summary
His tenacity was rewarded in 1977, when the Huskies won their conference championship and met the University of Michigan in the 1978 Rose Bowl game. The underdog Huskies won the Rose Bowl under Moon's leadership, and he was named Rose Bowl Most Valuable Player and the Pacific-8 Player of the Year. Overall, Moon passed for 3,277 yards and 19 touchdowns in his collegiate career.

Although Moon managed to win over Washington's fans, he failed to convince skeptical NFL scouts of his playing ability. His Rose Bowl performance notwithstanding, he was rated just the tenth best quarterback in the 1978 draft. "The stereotype was that he was a black quarterback and he was going to run around like a madman, but he wouldn't be able to throw very well," former Edmonton Eskimos and Houston Oilers coach Hugh Campbell told the *Los Angeles Times*. So, once again, Moon decided to prove himself elsewhere, signing with the Eskimos of the Canadian Football League.

During Moon's six seasons in Canada, he put up some stunning numbers—21,228 yards passing and 1,700 yards rushing. He had back-to-back 5,000-yard passing seasons. His 5,948 yards passing over 16 games in 1993 remains an all-time high for pro football. In addition, the Eskimos won five straight Grey Cup trophies as champions of the CFL from 1978 to 1982.

By 1984 Moon had nothing left to prove. When his contract with Edmonton expired, seven NFL teams sought to sign him as a free agent. Moon initially leaned toward the Seattle Seahawks, which would allow him to return to his college town, but he eventually chose the Houston Oilers, the team that had hired his former Edmonton coach, Campbell. The Oilers tendered a five-year, $5.5 million contract which, at the time, made Moon the highest paid player in the NFL—before he even played in a league game.

When Moon joined Houston, it was the20 worst franchise in the NFL, having won only three games in the previous two seasons. "One of the challenges of Houston was to be part of a growing situation," Moon's agent, Leigh Steinberg, told the *Houston Post*. "He knew it would take longer [to be on a championship team], but when he came he knew he would be an instrumental part of the building process.

In 1984 Moon was a rookie sensation. His six years in the CFL gave him a wealth of experience, and he threw for a then-Houston-record 3,338 yards on the season. Still the Oilers went 3-13, finishing last in their division. The next season, after the club won just five of its first 14 games, Campbell was fired and a defensive-oriented coach, Jerry Glanville, took over. "Those early years [in Houston] were really hard for me to deal with at first," Moon told the *St. Louis Post-Dispatch*. "There were some uncertainties about my career here because of the coaching change. That left me disheartened... Plus, I think most of the people looked at the amount of money I was paid and just decided I must be a star all of a sudden. I didn't respond well to it."
Short Answers

Answers 1-5

- Agent Leigh Steinberg
- Manny Ramirez Will Clark Steve
- Quarterback Warren
- Clark Steve Young
- Young Warren
What is Question Answering?

Goal: answer questions like “Who is Warren Moon’s Agent?”

Internet

10^6 webpage

Document Retrieval

Sentence Retrieval

Information Extraction

Answer Selection

1 Answer

e.g. “Leigh Steinberg”
Sentence Retrieval

- Task:
  - Finding a small segment of text that contains the answer
    [Corrada-Emmanuel, Croft, & Murdock, 2003]

- Benefits beyond document retrieval:
  - Documents are very large
  - Documents span different subject areas
  - The relevant information is expressed much more locally
  - Retrieving the sentences simplifies the next step:
    information extraction
Outline

✓ Question Answering System

❖ Language Models for Information Retrieval

☐ Trained Triggering Model

☐ Results

☐ Summary
Language Models for IR

- $P(Q|S_i)$: language model trained on $S_i$
- Ranking sentences in descending order of this probability
Query Likelihood Model

- Unigram language model of sentences and queries
  
  \[ P(Q \mid S) = \prod_{i=1}^{M} P(q_i \mid S) \]

[Song & Croft, 1999]
**Maximum Likelihood Estimation**

\[
P(Q | S) = \prod_{i=1}^{M} P(q_i | S)
\]

- A separate language model is trained for each sentence in the search space.
- The probability is calculated based on the frequency of query term in the sentence.

\[
P(q_i | S) = \frac{f_S(q_i)}{\sum_w f_S(w)}
\]
Example

Question: Who invented the automobile?

Answers:

An automobile powered by his own engine was invented by Karl Benz in 1885 and granted a patent.
Example

Question:
Who *invented* the *automobile*?

Answers:

An *automobile* powered by his own engine was *invented* by Karl Benz in 1885 and granted a patent.

Nicolas Joseph Cugnot invented the first self propelled mechanical vehicle.
Example

Question:

Who invented the automobile?

Answers:

- An automobile powered by his own engine was invented by Karl Benz in 1885 and granted a patent.
- Thomas Edison invented the first commercially practical light.
- Alexander Graham Bell of Scotland is the inventor of the first practical telephone.
Outline

✓ Question Answering System

✓ Language Models for Information Retrieval

❖ Trained Triggering Model

☐ Results

☐ Summary
**Trained Trigger Model**

\[
P(Q \mid S) = \prod_{i=1}^{M} P(q_i \mid S)
\]

- A unique model is trained on a large corpus firstly, then it is being used for all of the sentences to be retrieved.
- The trained model is represented by a set of triples:

  \[
  < w, w', f_C(w, w') >
  \]

- \( f_C(w, w') \) is the number of times the word \( w \) triggers the target word \( w' \).
Trained Trigger Model

\[ P(Q \mid S) = \prod_{i=1}^{M} P(q_i \mid S) \]

- Triggering Model:

\[ P(q_i \mid s_j) = \frac{f_c(q_i, s_j)}{\sum_{q_i} f_c(q_i, s_j)} \]

\[ P(q_i \mid S) = \frac{1}{N} \sum_{j=1}^{N} P(q_i \mid s_j) \]
Word Unigram Model

Query $Q$

$\theta_{S_2}$

$P(Q|S_2)$
Word Unigram Model

Query Q

\[ \theta_{S_5} \]

\[ P(Q|S_5) \]
Triggering Model

Query $Q$

$\theta_C$

$P(Q|S_2)$

Corpus $C$

$S_1, S_2, S_3, S_4, S_5, S_6, S_7$
Triggering Model

Query \( Q \)

\( \theta_C \)

\( P(Q|S_5) \)

Corpus C

\( S_1 \)

\( S_2 \)

\( S_3 \)

\( S_4 \)

\( S_5 \)

\( S_6 \)

\( S_7 \)
Word Unigram Model

Query

$n$-gram

$P(Q|S_2)$

$S_1$, $S_2$, $S_3$, $S_4$, $S_5$, $S_6$, $S_7$
Triggering Model

Query $Q$

Self Triggering
Inside Sentence
Across Sentence
Question and Answer Pair

$P(Q|S_2)$

$S_1$, $S_2$, $S_3$, $S_4$, $S_5$, $S_6$, $S_7$
**Self Triggering**

- Each word can only trigger itself

- Works similar to the basic bag-of-words model

- The words that appeared in both the query and the sentence have the higher priority

- It is an essential part of a retrieval engine and have to be used beside any other triggering models
Inside Sentence Triggering

- Idea: there is a relation between words appear in the same sentences
- Considers that each word in a sentence triggers all other words in the same sentence
- Uses a large unannotated corpus for training
- The sentence retrieval can retrieve sentences which do not share many terms with the query, but their terms frequently co-occur with query terms in the same sentences of the training corpus.
Example

Question: Who invented the automobile?

Answers:

Nicolas Joseph Cugnot invented the first self propelled mechanical vehicle.

Thomas Edison invented the first commercially practical light.

Alexander Graham Bell of Scotland is the inventor of the first practical telephone.
Inside Sentence Triggering

“The word \textbf{automobile} meaning a \textbf{vehicle} that moves itself.”

“An \textbf{automobile} includes at least two seats located one behind the other and attachable to a \textbf{vehicle} floor.”
Across Sentence Triggering

- Idea: two adjacent sentences mostly talk about the same topic by using different words with the same concept and meaning.
- Considers that each word of a sentence triggers all of the words of the next sentence in the training corpus.
- Uses a large unannotated corpus for trainings.
- Uses a wider context than inside sentence triggering.
Example

Question:

Where was the first McDonald’s built?

Answers:

Two brothers from Manchester, Dick and Mac McDonald, open the first McDonald’s in California.
Example

Question:

Where was the first McDonald’s built?

Answers:

Two brothers from Manchester, Dick and Mac McDonald, open the first McDonald’s in California.

The site of the first McDonald’s to be franchised by Ray Kroc is now a museum in Des Plaines, Illinois.

The first McDonald’s TV commercial was a pretty low-budget affair.
**Example**

**Question:**
Where was the *first McDonald’s* built?

**Answers:**
- Two brothers from Manchester, Dick and Mac McDonald, opened the *first McDonald’s* in California.
- The site of the *first McDonald’s* to be franchised by Ray Kroc is now a museum in Des Plaines, Illinois.
- The *first McDonald’s* TV commercial was a pretty low-budget affair.
Across Sentence Triggering

“The structure of Eiffel Tower was built between 1887 and 1889 as the entrance arch for the Exposition Universelle, a World’s Fair marking the centennial celebration of the French Revolution.”

“The tower was inaugurated on 31 March 1889, and opened on 6 May.”

“Wembley Stadium was built by Australian company Brookfield Multiplex.”

“The stadium was scheduled to open on 13 May 2006.”
Question and Answer Pair Triggering

- Idea: there is a relation between words appear in a pair of question and answer sentence

- Considers that each word in the question triggers all words in its answer sentence

- Requires a supervised training

- Uses a question-answer sentence corpus for training
Example

Question: How high is Everest?

Answers: Everest is 29,029 feet.
Example

Question: How high is Everest?

Answers:

- Everest is located in Nepal.
- Everest has two main climbing routes.
- Everest is 29,029 feet.

Note: The correct answer is Everest is 29,029 feet.
Example

Question: How high is Everest?

Answers:

- Everest is located in Nepal. (Wrong)
- Everest is 29,029 feet. (Correct)
- Everest has two main climbing routes. (Wrong)
Q: “How high is Pikes peak?”
A: “Pikes peak, Colorado At 14,110 feet, altitude sickness is a consideration when driving up this mountain.”

Q: “How high is Mount Hood?”
A: “Mount Hood is in the Cascade Mountain range and is 11,245 feet.”
Outline

✓ Question Answering System

✓ Language Models for Information Retrieval

✓ Trained Triggering Model

❖ Results

☐ Summary
Corpora

- **Inside Sentence**
- **Across Sentence**

**AQUAINT1 Corpus**
- Consists of English newswire text, extracted from:
  - the Xinhua (XIN)
  - the New York Times (NYT)
  - the Associated Press Worldstream (APW)
- Contains
  - almost 450 million word tokens
  - as 23 million sentences
  - as 1.5 million documents.

**Question and Answer Pair**

- **QASP corpus**
  - Consists of TREC QA track
    - from 2002 to 2004
    - 985 questions
  - Prepared via Amazon MTurk

- **Yahoo! Answers Corpus**
  - derived from
    - http://answers.yahoo.com/
  - collected in 10/25/2007
  - containing 4,483,032 questions
Experiments

- TREC data set
  - Development set: TREC05 (316 questions)
  - Test set: TREC06 (365 questions)

- Judgment
  - QASP Corpus
  - Prepared via Amazon MTurk

- Evaluation Metrics
  - Mean Average Precision
  - Precision@5
  - Mean Reciprocal Rank
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>MAP</th>
<th>MRR</th>
<th>P@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Likelihood</td>
<td>0.371</td>
<td>0.504</td>
<td>0.227</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>MAP</th>
<th>MRR</th>
<th>P@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Likelihood</td>
<td>0.3701</td>
<td>0.5047</td>
<td>0.2267</td>
</tr>
<tr>
<td>Self TTLM</td>
<td>0.3806</td>
<td>0.5219</td>
<td>0.2349</td>
</tr>
</tbody>
</table>
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>MAP</th>
<th>MRR</th>
<th>P@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Likelihood</td>
<td>0.3701</td>
<td>0.5047</td>
<td>0.2267</td>
</tr>
<tr>
<td>Self TTLM</td>
<td>0.3806</td>
<td>0.5219</td>
<td>0.2349</td>
</tr>
<tr>
<td>Inside Sentence TTLM</td>
<td>0.1911</td>
<td>0.2585</td>
<td>0.1052</td>
</tr>
<tr>
<td>Across Sentence TTLM</td>
<td>0.2367</td>
<td>0.3047</td>
<td>0.1360</td>
</tr>
<tr>
<td>Question and Answer Pair TTLM (QASP)</td>
<td>0.2266</td>
<td>0.3263</td>
<td>0.1360</td>
</tr>
<tr>
<td>Question and Answer Pair TTLM (Yahoo)</td>
<td>0.0344</td>
<td>0.0415</td>
<td>0.0099</td>
</tr>
</tbody>
</table>
**Results (Linear Interpolation)**

<table>
<thead>
<tr>
<th>Model</th>
<th>Self Triggering</th>
<th>Maximum Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAP</td>
<td>MRR</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.3806</td>
<td>0.5219</td>
</tr>
<tr>
<td>+ Inside Sentence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ Across Sentence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ QA Pair (QASP)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ QA Pair (Yahoo)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Results (Linear Interpolation)

<table>
<thead>
<tr>
<th>Model</th>
<th>Self Triggering</th>
<th>Maximum Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAP</td>
<td>MRR</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.3806</td>
<td>0.5219</td>
</tr>
<tr>
<td>+ Inside Sentence</td>
<td>0.4314</td>
<td>0.5469</td>
</tr>
<tr>
<td>+ Across Sentence</td>
<td>0.4274</td>
<td>0.5480</td>
</tr>
<tr>
<td>+ QA Pair (QASP)</td>
<td>0.4204</td>
<td>0.5408</td>
</tr>
<tr>
<td>+ QA Pair (Yahoo)</td>
<td>0.4358</td>
<td>0.5618</td>
</tr>
</tbody>
</table>

All the differences are statistically significant at the level of p-value < 0.01 based on t-tests.
## Results (Linear Interpolation)

<table>
<thead>
<tr>
<th>Model</th>
<th>Self Triggering</th>
<th>Maximum Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAP</td>
<td>MRR</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.3806</td>
<td>0.5219</td>
</tr>
<tr>
<td>+ Inside Sentence</td>
<td>0.4314</td>
<td>0.5469</td>
</tr>
<tr>
<td>+ Across Sentence</td>
<td>0.4274</td>
<td>0.5480</td>
</tr>
<tr>
<td>+ QA Pair (QASP)</td>
<td>0.4204</td>
<td>0.5408</td>
</tr>
<tr>
<td>+ QA Pair (Yahoo)</td>
<td>0.4358</td>
<td>0.5618</td>
</tr>
</tbody>
</table>

All the differences are statistically significant at the level of p-value<0.01 based on t-tests.
## Results (Linear Interpolation)

<table>
<thead>
<tr>
<th>Model</th>
<th>Self Triggering</th>
<th>Maximum Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAP</td>
<td>MRR</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.3806</td>
<td>0.5219</td>
</tr>
<tr>
<td>+ Inside Sentence</td>
<td>0.4314</td>
<td>0.5469</td>
</tr>
<tr>
<td>+ Across Sentence</td>
<td>0.4274</td>
<td>0.5480</td>
</tr>
<tr>
<td>+ QA Pair (QASP)</td>
<td>0.4204</td>
<td>0.5408</td>
</tr>
<tr>
<td>+ QA Pair (Yahoo)</td>
<td>0.4358</td>
<td>0.5618</td>
</tr>
</tbody>
</table>

All the differences are statistically significant at the level of p-value < 0.01 based on t-tests.
## Results (Linear Interpolation)

<table>
<thead>
<tr>
<th>Model</th>
<th>Self Triggering</th>
<th>Maximum Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAP</td>
<td>MRR</td>
</tr>
<tr>
<td>Baseline</td>
<td>0.3806</td>
<td>0.5219</td>
</tr>
<tr>
<td>+ Inside Sentence</td>
<td>0.4314</td>
<td>0.5469</td>
</tr>
<tr>
<td>+ Across Sentence</td>
<td>0.4274</td>
<td>0.5480</td>
</tr>
<tr>
<td>+ QA Pair (QASP)</td>
<td>0.4204</td>
<td>0.5408</td>
</tr>
<tr>
<td>+ QA Pair (Yahoo)</td>
<td>0.4358</td>
<td>0.5618</td>
</tr>
</tbody>
</table>

All the differences are statistically significant at the level of p-value < 0.01 based on t-tests.
Results (Linear Interpolation)
Outline

✓ Question Answering System
✓ Language Models for Information Retrieval
✓ Trained Triggering Model
✓ Results
✓ Summary
Summary

- Introducing question answering systems
- The necessity of the sentence retrieval in a QA system
- Using language models for sentence retrieval
- Describing the current unigram model and its problems
- Proposing trained triggering model with different types:
  - self
  - inside sentence
  - across sentence
  - question and answer pair
- Linear interpolation of different models
References

Thanks for your attention!