A Large-Scale Dataset for Known-Item Question Performance Prediction

QPP++ 2023, 6. April 2023, Dublin, Ireland

Maik Fröbe
Eric Oliver Schmidt
Matthias Hagen

1 FRIEDRICH-SCHILLER-UNIVERSITÄT JENA
2 MARTIN-LUTHER-UNIVERSITÄT HALLE-WITTENBERG

webis.de
A Large-Scale Dataset for Known-Item Question Performance Prediction

Motivation

- Query performance prediction experiments run often on TREC benchmarks
- Many currently used TREC benchmarks focus on informational queries, e.g.:
  - Robust04, MS MARCO, …
A Large-Scale Dataset for Known-Item Question Performance Prediction

Motivation

- Query performance prediction experiments run often on TREC benchmarks
- Many currently used TREC benchmarks focus on informational queries, e.g.:
  - Robust04, MS MARCO, …
A Large-Scale Dataset for Known-Item Question Performance Prediction

Motivation

- Query performance prediction experiments run often on TREC benchmarks
- Many currently used TREC benchmarks focus on informational queries, e.g.:
  - Robust04, MS MARCO, ...

- Many queries submitted to search engines are navigational [Broder'02]
- Known-item and re-finding queries, can be challenging [Hagen'15, Arguello'21]
  - Searchers might be unable to recall suitable identifier for the item
Motivation

Me, trying to recall the name of a movie...

It is on the tip of my tongue.
[Movie] From the 1990s?

I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who knows she was lying.
[Movie] From the 1990s?

I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who knows she was lying.

TREC 2023 Tip-of-the-Tongue Track https://trec-tot.github.io/
[Arguello'21, Bhargav'22]

The phenomenon of failing to retrieve something from memory, combined with partial recall and the feeling that retrieval is imminent.
A Large-Scale Dataset for Known-Item Question Performance Prediction

I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who knows she was lying.

TREC 2023 Tip-of-the-Tongue Track https://trec-tot.github.io/
[Arguello'21, Bhargav'22]

The phenomenon of failing to retrieve something from memory, combined with partial recall and the feeling that retrieval is imminent.

Tip-of-the-tongue Known-Item Retrieval Dataset (TOT-KIR):
[Bhargav'22]

- 15,000 questions with linked known items
- Focus: Movies and books
- Extracted from reddit.com/r/tipofmytongue/
- Crawled: 2017—2020
A Large-Scale Dataset for Known-Item Question Performance Prediction

Goal: Find the known-item

[Movie] From the 1990s?
I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who knows she was lying.

7 comments
Add a Comment

SassyPants5 • 53 Seconds later
Poison Ivy

sonz82 • 5 Minutes later
Wow that was fast, yes this is the one! Thank you!
Solved!

WhatIsThisBot • 2 Days later
You have been given one point for this answer. Thanks for contributing!
sonz82 awarded to SassyPants5 (6)

© Webis 2023
A Large-Scale Dataset for Known-Item Question Performance Prediction

Dataset construction

Assumption:

The time elapsed to answer a known-item question indicates its “performance”
A Large-Scale Dataset for Known-Item Question Performance Prediction

Dataset construction

Assumption:

The time elapsed to answer a known-item question indicates its “performance”

Example:

[Movie] From the 1990s?
I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who knows she was lying.

7 comments
Add a Comment

SassyPlants5 57 • 33 Seconds later
Poison Ivy
7 0 0 Reply

Sori82 510 • 6 Minutes later
Wow that was fast, yes this is the one! Thank you!
Solved!
A Large-Scale Dataset for Known-Item Question Performance Prediction

Dataset construction

Assumption:

The time elapsed to answer a known-item question indicates its “performance”

Example:

[Movie] From the 1990s?
I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who knew she was lying.

[late-2000s] Video clip of Thom Yorke talking about Radiohead album
Around the time “In Rainbows” came out, I remember seeing a video clip of Thom Yorke talking to the viewer. He did something weird and offbeat immediately, said something weird, maybe a little dance or moving his head around or something. Thanks in advance!
A Large-Scale Dataset for Known-Item Question Performance Prediction

Dataset construction

Assumption:

The time elapsed to answer a known-item question indicates its “performance”

Example:

[Movie] From the 1990s?
I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and lying that she was the passenger to the married man who she was lying.

7 comments
Add a Comment

[ late-2000s] Video clip of Thom Yorke talking about Radiohead album
Around the time “In Rainbows” came out, I remember seeing a video clip of Thom Yorke talking to the viewer. He did something weird and offbeat immediately, said something weird, maybe a little dance or moving his head around or something. Thanks in advance!

6 comments
Add a Comment

[Video] Recreation of Greys Anatomy surgery scene
These people recreated a scene in Greys Anatomy where someone puts on the surgeon’s gloves for them, then they operate and the patient dies and they continue to try and revive them. The caption is something along the lines of “Greys Anatomy be like...”

1 comments
Add a Comment

Not solved!
A Large-Scale Dataset for Known-Item Question Performance Prediction

Dataset construction

Assumption:

The time elapsed to answer a known-item question indicates its “performance”

Example:

[Movie] From the 1990s?
I remember a movie where a young woman having an affair with a married man in his 50s. The only scene I remember is the girl driving a car, crashing it, and laughing that she was the passenger to the married man who knew she was lying.

Our assumption allows assessing the performance of a question without requiring a link to the known item

Observation:

not solved!
A Large-Scale Dataset for Known-Item Question Performance Prediction

Dataset construction

- We crawl all questions and comments from reddit.com/r/tipofmytongue/
- 1.28 million questions
- Crawled: 2009—2022
- 47% are solved
- Many categories
  - Top-1 category: song with 160,000 questions
  - Top-2 category: movies with 159,000 questions
  - 13 categories with more than 10,000 questions
A Large-Scale Dataset for Known-Item Question Performance Prediction

Dataset construction

- We crawl all questions and comments from reddit.com/r/tipofmytongue/
- 1.28 million questions
- Crawled: 2009—2022
- 47% are solved
- Many categories
  - Top-1 category: song with 160,000 questions
  - Top-2 category: movies with 159,000 questions
  - 13 categories with more than 10,000 questions

Example data entry (jsonl format):

<table>
<thead>
<tr>
<th>Title</th>
<th>Question</th>
<th>UTC</th>
<th>Solved</th>
<th>Answer</th>
<th>Content</th>
<th>UTC</th>
</tr>
</thead>
</table>
A Large-Scale Dataset for Known-Item Question Performance Prediction

Pilot study: Known-Item Question Performance Prediction

- We use the qpptk toolkit\(^1\)
- We apply all seven pre-retrieval query performance predictors to our dataset

\(^1\)https://github.com/Zendelo/QPP-EnhancedEval/tree/main/code/qpptk
A Large-Scale Dataset for Known-Item Question Performance Prediction

Pilot study: Known-Item Question Performance Prediction

- We use the qpptk toolkit\(^2\)
- We apply all seven pre-retrieval query performance predictors to our dataset

<table>
<thead>
<tr>
<th></th>
<th>max-idf</th>
<th>avg-idf</th>
<th>scq</th>
<th>avg-scq</th>
<th>var</th>
<th>max-var</th>
<th>avg-var</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\tau)</td>
<td>.06</td>
<td>.06</td>
<td>.07</td>
<td>.02</td>
<td>.02</td>
<td>.11</td>
<td>.16</td>
</tr>
<tr>
<td>(\rho)</td>
<td>.02</td>
<td>.02</td>
<td>.07</td>
<td>.11</td>
<td>.16</td>
<td>.09</td>
<td>.05</td>
</tr>
<tr>
<td>(r)</td>
<td>.07</td>
<td>.07</td>
<td>.09</td>
<td>.09</td>
<td>.06</td>
<td>.02</td>
<td>.02</td>
</tr>
</tbody>
</table>

A Large-Scale Dataset for Known-Item Question Performance Prediction

Pilot study: Known-Item Question Performance Prediction

- We use the qpptk toolkit
- We apply all seven pre-retrieval query performance predictors to our dataset

<table>
<thead>
<tr>
<th>max-idf</th>
<th>avg-idf</th>
<th>scq</th>
<th>avg-scq</th>
<th>var</th>
<th>max-var</th>
<th>avg-var</th>
</tr>
</thead>
<tbody>
<tr>
<td>τ</td>
<td>ρ</td>
<td>r</td>
<td>τ</td>
<td>ρ</td>
<td>r</td>
<td>τ</td>
</tr>
<tr>
<td>.06</td>
<td>.06</td>
<td>.07</td>
<td>.02</td>
<td>.02</td>
<td>.07</td>
<td>.11</td>
</tr>
</tbody>
</table>

All predictors fail!

A Large-Scale Dataset for Known-Item Question Performance Prediction

Conclusion

- We presented a large dataset for known-item question performance prediction
- Proxy for the performance of a question = time elapsed until question is solved
- Pilot study on pre-retrieval predictors
  - None could really predict a known-item question’s performance.
- Data and code available:
  - https://github.com/webis-de/QPP-23
A Large-Scale Dataset for Known-Item Question Performance Prediction

Conclusion

- We presented a large dataset for known-item question performance prediction
- Proxy for the performance of a question = time elapsed until question is solved
- Pilot study on pre-retrieval predictors
  - None could really predict a known-item question’s performance.
- Data and code available:
  - https://github.com/webis-de/QPP-23

Future Work

- Ensure that there is no overlap to the TREC test topics
- Simplify data wrangling via ir_datasets
- Apply BERT-style predictors

We hope that the dataset supports the upcoming tip-of-my-tongue known-item retrieval track at TREC 2023
A Large-Scale Dataset for Known-Item Question Performance Prediction

Conclusion

- We presented a large dataset for known-item question performance prediction
- Proxy for the performance of a question = time elapsed until question is solved
- Pilot study on pre-retrieval predictors
  - None could really predict a known-item question’s performance.
- Data and code available:
  - https://github.com/webis-de/QPP-23

Future Work

- Ensure that there is no overlap to the TREC test topics
- Simplify data wrangling via ir_datasets
- Apply BERT-style predictors

We hope that the dataset supports the upcoming tip-of-my-tongue known-item retrieval track at TREC 2023

Thank You!