Team Bruce Banner at Touché 2022: Argument retrieval for controversial questions

Team: Bernardo Moreira, Bruno Martins, Fábio Goularte, and Henrique Cardoso

8th september 2022
Overview

- Introduction
- Background
- Problem
- Solution
- Experiments
- Conclusions
Argument retrieval focuses on retrieving the most pertinent information, given a query by a user;

Controversial topics are often sensitive topics;

To defend a point of view, one needs access to pertinent and reliable information;
The Touché task was presented at CLEF and focuses on retrieving arguments for controversial topics.
Goals

- Despite the focus for the 2022 edition of the Touché Task 1 being different from the 2021 edition, can a **simple IR approach** based on the construction of documents as **sentence pairs** obtain a competitive model for the Touché Task 1?
Contributions

○ With an IR approach based on sentence pairs, we achieved competitive results for the 2022 edition of the Touché Task 1;

○ Submitted a paper for the CLEF conference thoroughly describing our approach.
Approaches for the Touché Task 1

- **BM25**: Estimate relevance of documents to an input query.
- **TF-IDF**: Measure used to evaluate the relevance of a word in a document.
- **DirichletLM**: The Dirichlet function avoids unrecognized terms having estimated zero probability.
- **WordNet**: Used for query expansion.
Pyserini

- Python toolkit for information retrieval search;
- Sparse search using **Lucene** search library;
- Dense search using **Faiss** search library;
- Hybrid search using **HybridSearcher** which combines **Lucene** and **Faiss**.
Pyserini sparse search

Indexing

LuceneSearcher

BM25
N hits

Search

Top N Hits

Json File(s)

Query
For the 2022 edition of the Touché Task 1, the goal is to retrieve relevant pairs of sentences for a controversial question.
**Touché Task 1 2022 - Data**

**Id**
- "Sf9294c83-Af186e851"

**Conclusion**
- "Saddam Hussein is gone and Iraq is now functioning as one of very few democracies in the Middle East"

**Text**
- "His removal provides stability and security not only for Iraq but for the Middle East as a region"

**Stance**
- "PRO"

**discussionTitle/topic**
- "the War in Iraq was Worth the Cost"

**Sentences**
- 'sent_id': 'Sf9294c83-Af186e851__PREMISE__1',
  'sent_text': 'His removal provides stability and security not only for Iraq but for the Middle East as a region'
- 'sent_id': 'Sf9294c83-Af186e851__CONC__1',
  'sent_text': 'Saddam Hussein is gone and Iraq is now functioning as one of very few democracies in the Middle East'
Touché Task 1 2022 - Approach

Pair of sentences formation
Creates all the possible pairs from sentences in the same argument;

Queries and pairs variations
Create two types of queries and two types of pairs;

Sparse search
With the four different data combinations, Pyserini performs the sparse search.
It's important to be clear that this debate is looking at the results of the Iraq war and, by any definition Iraq is in a much more stable and secure position than it was in 2003 when American, British and other international troops arrived in the country.

Whatever one thinks of the initial justifications for the war there is no doubt that the country, the region and the world are better and safer places without Saddam Hussein.

Saddam Hussein is gone and Iraq is now functioning as one of very few democracies in the Middle East.
The state often takes years to recognise the needs of students and they lose years of education in the process. In addition, even if those needs are identified ‘special schools’ are underfunded and stigmatised.

Sentence pair with topic

The state often takes years to recognise the needs of students and they lose years of education in the process. In addition, even if those needs are identified ‘special schools’ are underfunded and stigmatised. Home schooling
<topic><number>1</number><title>Should teachers get tenure?</title> Query with title only
<description>A user has heard that some countries do give teachers tenure and others don’t. Interested in the reasoning for or against tenure, the user searches for positive and negative arguments. The situation of school teachers vs. university professors is of interest. </description>
narrative>Highly relevant arguments make a clear statement about tenure for teachers in schools or universities. Relevant arguments consider tenure more generally, not specifically for teachers, or, instead of talking about tenure, consider the situation of teachers’ financial independence. </narrative>
</topic>

Query with title, description and narrative concatenated;
## Touché Task 1 2022 - Runs

<table>
<thead>
<tr>
<th></th>
<th>Query title</th>
<th>Query title, description, and narrative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence pair</td>
<td>pyserini_sparse_v2</td>
<td>pyserini_sparse_v1</td>
</tr>
<tr>
<td>Sentence pair with topic</td>
<td>pyserini_sparse_v4</td>
<td>pyserini_sparse_v3</td>
</tr>
</tbody>
</table>
Experiments - Touché Task 1 2022

- Touché provided the evaluation files;

- Human assessors label the retrieved documents manually;

- The metrics used for evaluation are:
  - rhetorical quality of the document;
  - general topic relevance;
  - whether the sentence pair forms a coherent text;

- Trec_eval was used to compare our results with the evaluation files.
## Touché Task 1 2022 - Submitted runs results

<table>
<thead>
<tr>
<th>Run</th>
<th>Quality NDCG@5</th>
<th>Relevance NDCG@5</th>
<th>Coherence NDCG@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>pyserini_sparse_v1</td>
<td>0.772</td>
<td>0.641</td>
<td>0.378</td>
</tr>
<tr>
<td>pyserini_sparse_v2</td>
<td>0.701</td>
<td>0.580</td>
<td>0.353</td>
</tr>
<tr>
<td>pyserini_sparse_v3</td>
<td>0.760</td>
<td>0.651</td>
<td>0.354</td>
</tr>
<tr>
<td>pyserini_sparse_v4</td>
<td>0.709</td>
<td>0.586</td>
<td>0.357</td>
</tr>
</tbody>
</table>
# Touché Task 1 2022 - Classification

<table>
<thead>
<tr>
<th>Team</th>
<th>Relevance NDCG@5</th>
<th>Team</th>
<th>Quality NDCG@5</th>
<th>Team</th>
<th>Coherence NDCG@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porthos</td>
<td>0.742</td>
<td>Daario Naharis</td>
<td>0.913</td>
<td>Daario Naharis</td>
<td>0.458</td>
</tr>
<tr>
<td>Daario Naharis</td>
<td>0.683</td>
<td>Porthos</td>
<td>0.873</td>
<td>Porthos</td>
<td>0.429</td>
</tr>
<tr>
<td><strong>Bruce Banner</strong></td>
<td>0.651</td>
<td>Gamora</td>
<td>0.785</td>
<td>Gamora</td>
<td>0.398</td>
</tr>
<tr>
<td>D Artagnan</td>
<td>0.642</td>
<td>Hit Girl</td>
<td>0.776</td>
<td>Hit Girl</td>
<td><strong>0.378</strong></td>
</tr>
<tr>
<td>Gamora</td>
<td>0.616</td>
<td>Gorgon</td>
<td>0.742</td>
<td>D Artagnan</td>
<td>0.378</td>
</tr>
<tr>
<td>Hit Girl</td>
<td>0.588</td>
<td>D Artagnan</td>
<td>0.733</td>
<td>Hit Girl</td>
<td>0.377</td>
</tr>
<tr>
<td>Pearl</td>
<td>0.481</td>
<td>Pearl</td>
<td>0.678</td>
<td>Gamora</td>
<td>0.285</td>
</tr>
<tr>
<td>Gorgon</td>
<td>0.408</td>
<td>Swordsman</td>
<td>0.608</td>
<td>Gorgon</td>
<td>0.282</td>
</tr>
<tr>
<td>General Grievous</td>
<td>0.403</td>
<td>General Grievous</td>
<td>0.517</td>
<td>Swordsman</td>
<td>0.248</td>
</tr>
<tr>
<td>Swordsman</td>
<td>0.356</td>
<td>Korg</td>
<td>0.453</td>
<td>General Grievous</td>
<td>0.231</td>
</tr>
<tr>
<td>Korg</td>
<td>0.252</td>
<td></td>
<td></td>
<td>Korg</td>
<td>0.168</td>
</tr>
</tbody>
</table>
Conclusion

For the 2022 edition of the Touché Task 1 the Pyserini sparse search resulted in third, fourth and fifth places in relevance, coherence and quality, respectively;

The results for the 2022 edition could be improved by using a different data arrangement.
Future work

- Different data arrangement to form sentence pairs;

- Develop a considerably smaller sentence-pairs dataset to use Pyserini hybrid search.
Thank you for your attention!