Overview of the Causality-driven Adhoc Information Retrieval (CAIR) task at FIRE-2021

Suchana Datta[1], Debasis Ganguly[2], Dwaipayan Roy[3] and Derek Greene[1]

[1]University College Dublin, Ireland
[2]University of Glasgow, United Kingdom
[3]Indian Institute of Science Education and Research, Kolkata

Abstract

The CAusality-based Information Retrieval (CAIR) track at FIRE 2021 focuses on the task of retrieving potentially relevant documents in response to a query indicating one or more events, where the notion of relevance is determined by whether a document indicates potential causes that might have led to the specified events in the query. In 2020, we released a dataset comprised of a benchmark set of 25 queries along with the relevance judgments. The target document collection is the English monolingual FIRE ad-hoc document collection. This second iteration of the track acted as a continuation of the same task with the same dataset as in the last year. The objective was to encourage the participants to try out more involved approaches (e.g. supervised ones) for improving on the retrieval effectiveness.

Keywords

Causal Information Retrieval, Semantic Search, BERT, Apache Nutch

1. Introduction

In traditional ad-hoc IR setup, a search system retrieves a ranked list of documents given a query. The usefulness of the output of an ad-hoc IR system, in the form of a ranked list of documents, is limited in situations when i) decision makers need to formulate policies to mitigate a current event that requires attention (e.g. drop in the value of British pound), or ii) policy-making regarding societal benefits (e.g. formulating government policies to reduce housing crisis by analyzing the main likely causes). In the aforementioned situations, a traditional search system user is required to carefully analyze the topically relevant documents (likely to describe the main event expressed in the query itself) and most likely needs to reformulate queries in order to retrieve documents related to the potential *causes leading* to the (query) event. The user of a traditional IR system, hence, needs to spend considerable effort in reformulating queries in order to retrieve the causally relevant documents towards top ranks [1, 2].

With this motivation, we proposed a shared task for the first time in FIRE 2020 [3] to investigate approaches to reduce this manual effort and ask participants to design effective

FIRE'21: Forum for Information Retrieval Evaluation, December 13–17, 2021, Virtual Event
suchana.datta@ucdconnect.ie (S. Datta); debasis.ganguly@glasgow.ac.uk (D. Ganguly); dwaipayan.roy@iiserkol.ac.in (D. Roy); derek.greene@ucd.ie (D. Greene)
https://csintranet.ucd.ie/phd-student/suchana-datta/ (S. Datta); https://gdebasis.github.io/ (D. Ganguly); https://dwaipayanroy.github.io/ (D. Roy); http://derekgreene.com/ (D. Greene)

CEUR Workshop Proceedings (CEUR-WS.org)

retrieval models seeking to address *causality-based relevance* [1, 2] rather than the traditional *topical relevance.*

We provided participants a static test collection of 303, 291 news documents and a list of 25 queries, divided into two parts - 5 queries for training and 20 queries for test, related to events that were likely to be caused by a number of other past events. We also provided associated relevance judgements for the set of train queries. The participants were to required to develop ranking models that could effectively retrieve documents containing information on such past events which were likely candidates to lead to the query event.

From the first iteration of CAIR track [3], the two main observations were that, firstly, longer queries showed a general trend to yield more causally relevant documents towards top ranks as seen from the results obtained from the first participating group [4]; and secondly, it also turned out that sequence-based text representation for semantically matching the documents with queries did not yield effective retrieval results [5] and thus leaving open a scope to apply more involved approaches for addressing the task of causality-based retrieval. With this motivation, we have continued the second iteration of CAIR task at FIRE 2021[1].

2. Dataset

As the notion of causality differs from the idea of topical relevance, the selection of topics for this task was restricted to the query events with causal information need as we detailed in [3]. In this section, we briefly recapitulate the dataset construction process.

Target Collection The second iteration of the track uses the same target collection as in the first year [3], i.e., the English ad-hoc IR collection of FIRE [6]. The news articles were crawled from the source 'Telegraph India'[2] published over a period of 10 years (2001 to 2011) [3].

Query Formulation This year the task was run with same query set as in 2020. As mentioned in [3], while selecting the topics, we took the following into consideration.

1. We ensured that a query is representative of an event that occurred during the period covered by the target collection, i.e. between 2001-2011.
2. An event qualifies as a valid topic only if there exists a multiple number of potential (arguable) causes that might have led to it. We eliminate those cases where the notion of causality is mentioned in the same document also describing the query event or it does not help user to walk through the chain of query event at all.

Relevance Assessments In contast to the previous year, where we extended the judgment pool with new documents obtained from the runs submitted by participating teams, this year we did not extend the relevance pool. This means that the relevance assessments are also identical to what was used in CAIR 2020 [3].

[1] http://fire.irsi.res.in/fire/2021/home
[2] https://www.telegraphindia.com/

3. Models Proposed by Participating Teams

We received a total of *four* submissions from *two* participating teams this year. The proposed model architectures are as follows:

NUIG [7] This team is comprised of participants both from National University of Ireland and University College Cork. The team proposed a semantic search pipeline that aggregates results across multiple query strategies and indices (a lexical and a semantic index). The authors applied three query strategies, denoted as Q_1, Q_2, and Q_3, where, Q_1 embeds the query using the sentence embedding model and retrieves the most relevant results based on cosine similarity; Q_2 and Q_3 retrieve the most relevant documents from the lexical index. Finally, Q_3 adds filtering and keyword extraction steps to transform the narrative description in causal search terms. In the end, results from all three queries ($\mathcal{D}_1, \mathcal{D}_2$, and $\mathcal{D}_3$ respectively) are aggregated and re-ranked by the aggregating module.

NITS [8] Team 'NITS' is from National Institute of Technology, Silchar. They investigated the potential of neural network-based language representation models, specifically the BERT model [9] and Apache Nutch [10] for the task of causal documents retrieval. The BERT model transformed the sequence of words into fixed-size embedding vector and used cosine similarity for relevance measurement; whereas, Apache Nutch is a keyword matching approach and used an AND search module to retrieve news articles that match to the input query. However, the results showed that sequence-based text representation for semantically matching the documents with queries did not yield effective retrieval results.

4. Evaluation

Each participating team was allowed to submit at most three runs. Team NUIG has submitted single run while three runs were submitted by team NITS. We evaluate each submitted run based on their performance achieved over 20 test queries. In particular, we used the following evaluating measures to report model's efficiency:

- **MAP:** We chose Mean Average Precision (MAP) as our primary measure of retrieval effectiveness so as to take both precision and recall into account. This metric quantifies the retrieval model based on the mean of the average precision scores achieved per query.
- **P@5:** We also made use of $P@5$ to measure model's efficacy, i.e. number of relevant documents present in the top 5 ranked documents and averaged over test query set.

The performances are evaluated using 'trec-eval' [3] and the results are reported in Table 1.

5. Results

Followings are the few observations that we made from this year submissions (see Table 1) :

[3]https://trec.nist.gov/trec eval/

- The best results this year are better than the results from 2020 [3]. It is worth mentioning that the results are comparable across the two versions of the track, because the dataset is identical.
- NUIG's supervised approach [7], pairwise training of causality specific similarity, outperforms an unsupervised approach [8].
- As claimed by authors in [1] which proposes an unsupervised technique, causally connected documents are likely to have only a partial term overlap with the corresponding topical set, query narrations are certainly a good resource of finding such causality specific terms given the query event. However, it turned out that retrieving only with titles is better than also making use of the narratives.
- Supervised approach turned out to be better than query enrichment methods of the last year.

Year	**Team**	**Run ID**	**MAP**	**P@5**	**Model Summary**
2021	NUIG	**semantic_IR**	**0.5761**	**0.7800**	semantic search
	NITS	BERT(Q_{title})	0.0014	0.0100	embedding, semantic similarity
		BERT(Q_{narr})	0.0001	0.0000	
		Nutch(Q_{title})	0.1063	0.4800	
2020	UCSC	Q_{narr}	**0.4553**	**0.7000**	detect causal relations, query expansion
		Q_{title}	0.4066	0.5400	
		post-event -terms-expan	0.3885	0.5000	
	NITS	run-1	0.0577	0.2600	embedding with USE [11]

Table 1
Retrieval effectiveness of models proposed by participating teams of both 2020 and 2021. Best performing model outcome by year is bold-faced and the performance of the topmost team so far is underlined.

References

[1] S. Datta, D. Ganguly, D. Roy, F. Bonin, C. Jochim, M. Mitra, Retrieving potential causes from a query event, in: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '20, Association for Computing Machinery, 2020, p. 1689–1692.

[2] S. Datta, D. Greene, D. Ganguly, D. Roy, M. Mitra, Where's the why? in search of chains of causes for query events, in: Proceedings of The 28th Irish Conference on Artificial Intelligence and Cognitive Science, Dublin, Republic of Ireland, December 7-8, 2020, volume 2771 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 109–120.

[3] S. Datta, D. Ganguly, D. Roy, D. Greene, C. Jochim, F. Bonin, Overview of the causality-driven adhoc information retrieval (CAIR) task at FIRE-2020, in: FIRE 2020: Forum for Information Retrieval Evaluation, Hyderabad, India, December 16-20, 2020, ACM, 2020, pp. 14–17.
[4] C. Lin, Y. Zhang, Causality detection for causality-driven adhoc information retrieval task, in: Proceedings of FIRE 2020 - Forum for Information Retrieval Evaluation (December 2020), 2020.
[5] P. Dadure, P. Pakray, S. Bandyopadhyay, Preliminary investigation on causality information retrieval, in: Proceedings of FIRE 2020 - Forum for Information Retrieval Evaluation (December 2020), 2020.
[6] S. Palchowdhury, P. Majumder, D. Pal, A. Bandyopadhyay, M. Mitra, Overview of FIRE 2011, in: Multilingual Information Access in South Asian Languages - Second International Workshop, FIRE 2010, Gandhinagar, India, February 19-21, 2010 and Third International Workshop, FIRE 2011, Bombay, India, December 2-4, 2011, Revised Selected Papers, 2011, pp. 1–12.
[7] D. Dalal, S. D. Gupta, B. Binaei, A semantic search pipeline for causality-driven adhoc information retrieval, in: Proceedings of FIRE 2021 - Forum for Information Retrieval Evaluation (December 2021), 2021.
[8] P. Dadure, P. Pakray, S. Bandyopadhyay, Causal document retrieval using bert and apache nutch, in: Proceedings of FIRE 2021 - Forum for Information Retrieval Evaluation (December 2021), 2021.
[9] J. Devlin, M. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2019. arXiv:1810.04805.
[10] R. Khare, R. C. Douglas, K. Sitaker, A. Rifkin, Nutch: A flexible and scalable open-source web search engine, 2005.
[11] D. Cer, Y. Yang, S. Kong, N. Hua, N. Limtiaco, R. S. John, N. Constant, M. Guajardo-Cespedes, S. Yuan, C. Tar, Y. Sung, B. Strope, R. Kurzweil, Universal sentence encoder, 2018. arXiv:1803.11175.