

CincyMedIR at TREC 2021 Clinical Trial Track

Hoang Vu^{1,2}, Danny T.Y. Wu, PhD, MSI^{1,2}

¹Department of Biomedical Informatics, University of Cincinnati College of Medicine, Cincinnati, OH; ²Department of Electrical Engineering and Computer Science, College of Engineering and Applied Science, University of Cincinnati, Cincinnati, OH

Introduction

The CincyMedIR team led by Dr. Danny T.Y. Wu at the University of Cincinnati College of Medicine participated in the Text Retrieval Conference 2021 Clinical Trials Track (TREC-CT). This year, we applied our existing text-retrieval methods from our previous TREC tracks on the new Clinical Trials track while also experimenting with new techniques to process the trial exclusion criteria.

Method

Similar to our past TREC tracks, we continued to use Elasticsearch as the information retrieval platform. However, for the nature of clinical trials, this time we focused more on filtering based on the extracted information, including age, gender, and exclusion criteria, prior to performing the text search. We applied three filters to the clinical trials. First of all, we extracted the patient's age and gender from the topic file by applying a list of regular expression rules and filtering only trials whose age and gender requirements match the patient's. Secondly, we filtered only trials with the "recruiting" status. Thirdly, we extracted negating concepts from both the topic description and the trial exclusion criteria and crossed out trials with matching negating concepts with the topic description. We used the python pipeline negspaCy paired with the python model scispaCy "en_ner_bc5cdr_md" to extract negating chemical and disease concepts from the texts. In the end, the top 1000 trials are retrieved for each topic by performing text search with Elasticsearch that queried the topic description against the trial title, brief summary, and detailed description. Our different runs are just different orderings of filters and search, i.e., which filters come before the search (pre-filters) and which filters come after the search (reranking). For our past TREC tracks, we also used Learning to Rank algorithms to sort the search results based on relevance feedback provided by TREC, but we did not do it this time as we did not find supervised data relevant to the Clinical Trials corpus.

Results

Run ID	Description	ncdg_cut_10	P_10	recip_rank
CincyMedIR_1	Pre-filters: age, gender, status, exclusion criteria Reranking: none	0.1744	0.1187	0.2348
CincyMedIR_2	Pre-filters: age, gender, status Reranking: exclusion criteria	0.1744	0.1187	0.2348

CincyMedIR_3	Pre-filters: age, gender, exclusion criteria Reranking: status	0.1744	0.1187	0.2356
CincyMedIR_4	Pre-filters: age, gender Reranking: status (first), exclusion criteria (second)	0.1744	0.1187	0.2357
CincyMedIR_5	Pre-filters: age, gender Reranking: exclusion criteria (first), status (second)	0.1744	0.1187	0.2360
CincyMedIR_6*	Pre-filters: none Reranking: none	0.1923	0.0973	0.2751
CincyMedIR_7*	Pre-filters: age, gender Reranking: none	0.2188	0.1267	0.3213
CincyMedIR_8*	Pre-filters: age, gender, status Reranking: none	0.1701	0.1133	0.2541
CincyMedIR_9*	Pre-filters: age, gender, exclusion criteria Reranking: none	0.2019	0.1253	0.3189
Median	Median of all automated runs	0.3040	0.1613	0.2942

Table 1. Performance of runs

* These runs were not submitted to TREC-CT but generated after the conference using trec_eval.

Table 1 shows that the performance of our submitted runs (1, 2, 3, 4, and 5) did not differ much. For ndcg_cut_10 and P_10, the results were exactly the same. For recip_rank, the results were slightly different. It can be inferred that the top 10 relevant trials retrieved by all runs were the same, which makes sense because the ordering of filters and search does not matter for the top few relevant trials. All our scores were below the median. The scores listed in table 1 are the means of the individual score for each topic. For some topics, the performance of our runs surpassed the median. The ID of these topics are: 1, 3, 4, 7, 11, 14, 15, 16, 20, 25, 26, 27, 31, 42, 44, 45, 51, 61, 65, and 73.

After the conference, we learned that our filters were counterproductive as they hard-filtered out many relevant trials. We generated four more runs where some filters are completely removed. Most of these additional runs showed better performance. Age and gender are the only pre-filters of the best run, CincyMedIR_7, which suggested that they were the only good hard filters. Recruiting status was a bad filter since it significantly lowered the scores as seen in

CincyMedIR_8. Exclusion criteria did not lower the scores as much; it provided some useful information, yet should not be used as a hard filter.

Discussion

We did not achieve high performance for the Clinical Trials track this year. The hard-filtering method dragged the performance. The additional runs that have all the filters removed except age and gender achieved better. While the recruiting status did not seem to be relevant, we have gathered useful information by extracting negation terms from the exclusion criteria. We plan to refine our method by turning the exclusion criteria matching from a hard filter into a score and combining it with the text matching score. We expect to utilize Learning to Rank algorithms using relevance feedback as the training data and continue to learn from other teams' techniques at the conference. Lastly, we thank the organizers of TREC-CT for their effort and look forward to participating in the track next year.