

Problem Description

We tackle the challenge of **ranking 450k clinical trials** based on **patient eligibility across 40 topics**. The queries are responses to a patient questionnaire based on a disorder.

Two Stage Clinical Trial Ranking Approach

We use a two-stage retrieval methodology:

Stage 1: Efficient ranking using normalized clinical concepts.

Stage 2: Fine-grained relevance scoring based on semantic similarity.

Stage 1: Rank

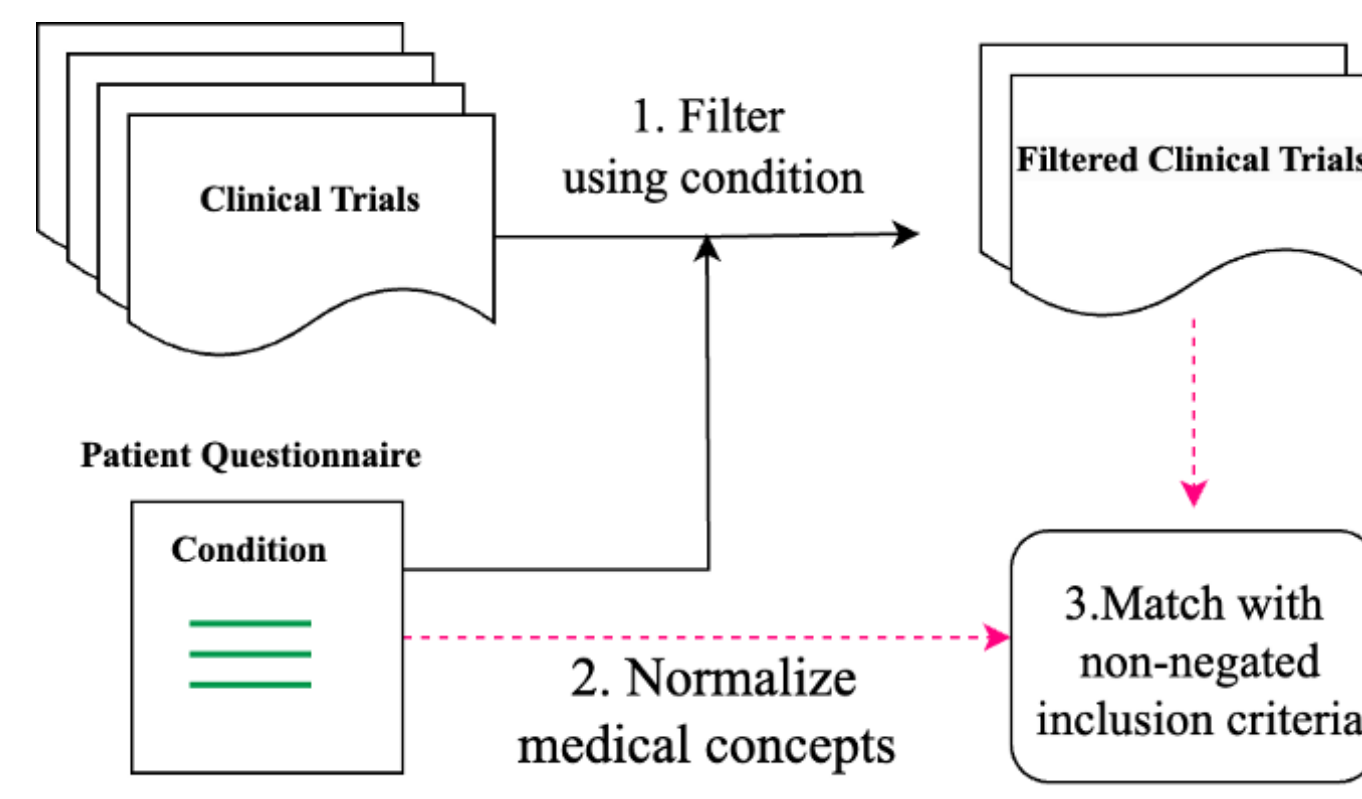


Fig. 1: Stage 1:

Step 1: Extract the inclusion and exclusion criteria and normalize the medical concepts using a knowledge base (UMLS).

Step 2: Match the concepts in the patient's questionnaire with the extracted concepts

Step 3: Rank based on the number of matches.

Stage 2: Re-rank

Step 1: Itemize the inclusion and exclusion criteria.

Step 2: Convert the questionnaire responses into sentences.

Step 3: Encode the itemized criteria and responses into contextual embeddings.

Step 4: Compute pairwise cosine similarity between criteria and the responses, resulting in two matrices.

Step 5: Compute ranking scores based on the similarity matrices.

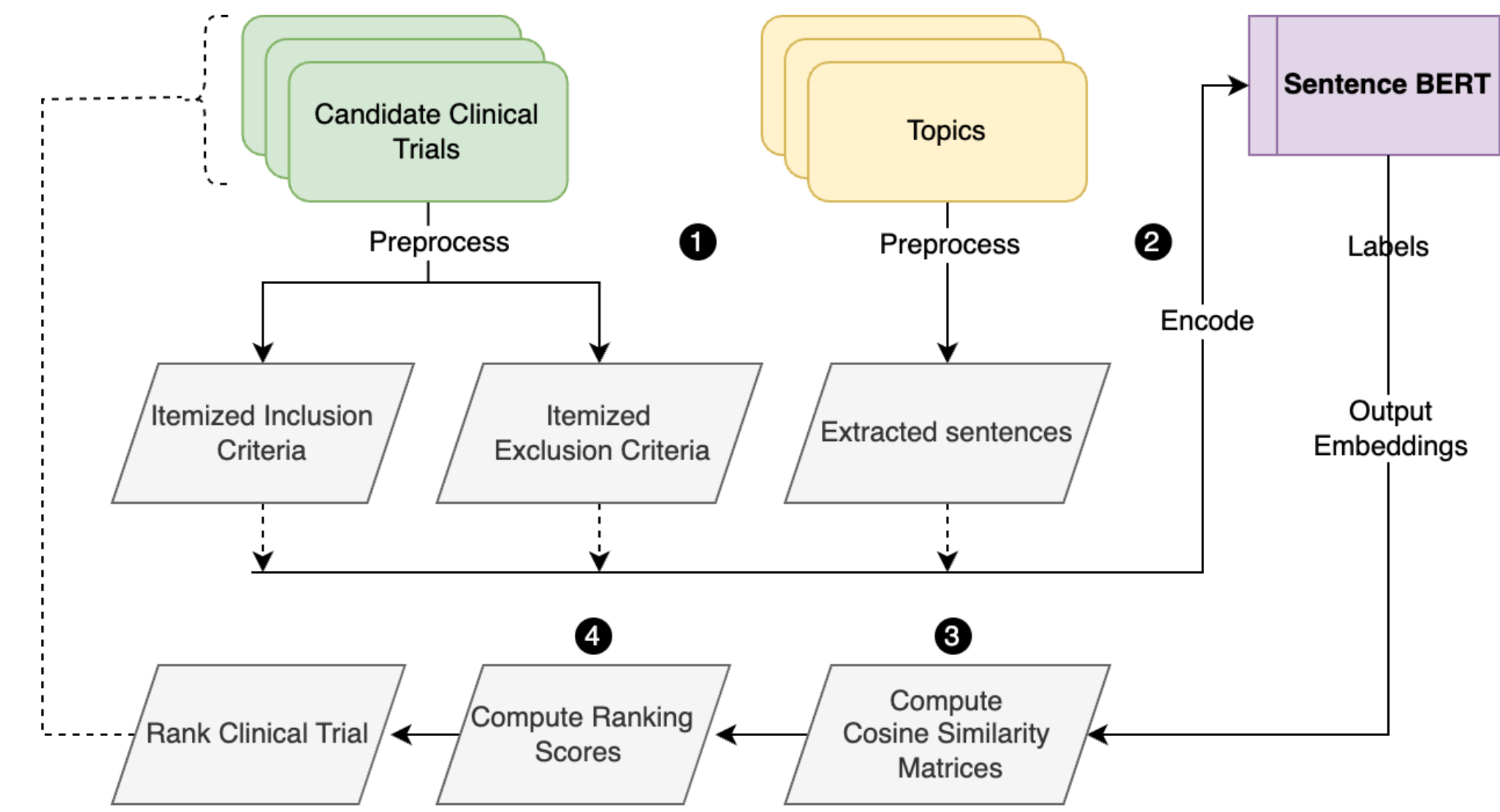


Fig. 2: Stage 2 Re-ranking Pipeline uses Cosine Similarity matrices for Criteria, Topic Sentences

Relevance score using similarity matrices

The relevance score is a function of two similarity matrices (shown in Fig. 3).

$$Score = Reward + Penalty$$

where,

- A positive Reward is given for matching an inclusion criteria.
- A negative Penalty is given for matching an exclusion criteria.

		Topic Sentences				
		Sent 1	Sent 2	Sent 3	Sent 4	Sent 5
Inclusion Criteria	Criteria 1	0.1	0.5	- 0.6	0.1	0.5
	Criteria 2	0.6	0.2	0.2	- 0.8	0.3
	Criteria 3	0.38	- 0.4	- 0.5	0.76	0.42
	Criteria 4	0.45	- 0.57	0.23	0.21	0.43
		<i>I</i>				

		Topic Sentences				
		Sent 1	Sent 2	Sent 3	Sent 4	Sent 5
Exclusion Criteria	Criteria 1	0.4	0.8	- 0.3	0.1	0.5
	Criteria 2	0.6	0.1	0.2	- 0.6	0.3
	Criteria 3	0.2	- 0.4	- 0.5	0.8	0.7
		<i>E</i>				

Fig. 3: Pairwise cosine similarity Matrices *I* and *E*.

Naïve High Precision: A full score is given if a patient matches all the inclusion criteria (at least once) And matches none of the exclusion criteria. A match is made using a threshold.

Weighted Relevance:

$$W_{Score} = Mean(I) - Max(E)$$

Matching exclusion criteria results in a higher penalty.

Balanced Relevance:

$$B_{Score} = Mean(I) - Mean(E)$$

Equal weights are given to both the reward and the penalty.

Experiments and Results

Runs	Precision@10 (min/median/max)	Mean Average Precision (min/median/max)	Reciprocal rank (min/median/max)
Stage1	0.00 / 0.5 / 1	0.00 / 0.07 / 0.26	0.00 / 1 / 1
Naïve relevance rank	0.00 / 0.5 / 1	0.00 / 0.08 / 0.31	0.00 / 0.5 / 1
Balanced Relevance Rank	0.00 / 0.70 / 1	0.00 / 0.14 / 0.36	0.00 / 1 / 1
Weighted Relevance Rank	0.00 / 0.60 / 1	0.00 / 0.12 / 0.36	0.00 / 1 / 1
Across all Participants (Averaged over all topics)	0.00 / 0.39 / 0.88	0.00 / 0.09 / 0.25	0.01 / 0.53 / 1.00

Fig. 4: Minimum, median, and maximum of Trec-eval metrics. All submissions are better than the median score of P-10. When considering the median, balanced relevance ranking outperforms other runs in Binary relevance metrics.

Runs	NDCG-cut-10 (min/median/max)	NDCG-cut-1000 (min/median/max)
Stage1	0.00 / 0.71 / 1	0.02 / 0.29 / 0.48
Naïve relevance rank	0.00 / 0.63 / 1	0.01 / 0.28 / 0.5
Balanced Relevance Rank	0.06 / 0.82 / 1	0.13 / 0.31 / 0.51
Weighted Relevance Rank	0.06 / 0.73 / 1	0.01 / 0.28 / 0.52
Across all Participants (Averaged over all topics)	0.01 / 0.64 / 0.92	0.00 / 0.39 / 0.54

Fig. 5: Balanced relevance ranking performs better than other runs if medians are considered. Two runs are better than the median NDCG-cut-10 scores.

Conclusions

We use a two-stage methodology of ranking and re-ranking. We propose three different strategies for computing relevance scores, based on similarity matrices.

- Hard matching using naïve high precision score.
- Weighted score to over-penalize exclusion matching.
- Balanced score for equal reward and penalty for matching inclusion and exclusion criteria.

Our submission performs better in terms of Precision@10 and NDCG-cut-10 than the median scores. Balanced relevance ranking outperforms the other approaches in terms of median metrics.