Query generation module annotates topics using IBM Watson Annotator for Clinical Data. Features specific to the topics were ranked using MIMIC, PubMed, and a rare lexicon disease to prioritize uncommon conditions.

Retrieval modules using a BM25 retrieval on a Lucene index and a transformer based semantic textual similarity (STS) model retrieve candidate trials.

Neural BERT-Based rerankers rank the top 2k trials per topic. See a diagram of our reranker architecture below. Entities are extracted from topics and trials using a model trained on CHIA Text is encoded using a BERT-Based model (CT-BERT), and attention mechanisms are used to compute alignment between spans in the topics and the criteria, interventions, MeSH terms, keywords, and condition.

With limited training data availability, we leverage information in both the MIMIC III corpus and clinical trials to create a large-scale silver-standard dataset.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Patient Description</th>
<th>Clinical Trials</th>
<th>Labeled Pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>TREC</td>
<td>75</td>
<td>375K</td>
<td>0</td>
</tr>
<tr>
<td>SIGIR</td>
<td>60</td>
<td>204K</td>
<td>3870</td>
</tr>
<tr>
<td>AutoGT</td>
<td>18k</td>
<td>375K</td>
<td>700K+</td>
</tr>
</tbody>
</table>

• Our base retrieval system (IBMLucene) achieved the best results across three metrics.
• Existing SOTA information retrieval methods led to poor performance on development data, and we did not submit runs using them.
• Large-scale training data for this problem is not available, and the proposed silver-standard training data approaches led to worse performance by our deep learning systems.
• We plan to perform deeper analysis to identify the causes once the relevance judgements are publicly released.

Citations