COL-764 Project Results Analysis
Precedent Retreival of Legal Cases

Pratik, Prawar, Rohan Debbarma,
Indian Institute of Techonology, Delhi
cs5180415@iitd.ac.in, cs5180417@iitd.ac.in

Abstract

This report presents analysis of aggregating each sub-query described in previous reports as a single ranking and then ranking. In addition, we also report retrieval time analysis of the methods performing the best for each category such as BM25, TF-IDF etc. In addition, KL divergence metric is also used in LDA based method to model query-document similarity. However, the results for KL divergence based retrieval were quite poor as compared to cosine similarity.

1 Introduction

This report describes some of the further analysis of the methods used in the precedent retrieval task.

2 Conjuctive Aggregation of Sub-queries vs Max-Pooling

In this section, we analyse aggregating each of the sub-queries as a single query and then ranking it with the original results.

Table 1: Scoring Approach Analysis of Methods

Method Used	Scoring Approach	MAP	MRR	$P@10$	$Rec@100$
BM-25	Conjunctive Queries	0.355	0.66	0.221	0.75
Doc2Vec	Conjunctive Queries	0.271	0.58	0.162	0.626
TF-IDF	Conjunctive Queries	0.33	0.62	0.2	0.73
LDA	Conjunctive Queries	0.102	0.234	0.07	0.51
Word2Vec	Conjunctive Queries	0.062	0.088	0.07	0.521
BM-25	Max-pooling Subqueries	0.482	0.823	0.277	0.794
Doc2Vec	Max-pooling Subqueries	0.386	0.747	0.229	0.705
TF-IDF	Max-pooling Subqueries	0.368	0.710	0.225	0.754
LDA	Max-pooling Subqueries	0.099	0.238	0.070	0.508
Word2Vec	Max-pooling Subqueries	0.083	0.111	0.093	0.609

2.1 Analysis of Results

From the above results, we can see that Max-pooling sub-queries gives better results for all the methods other than LDA whereas for LDA it give slightly better results with conjunction instead of max pooling.

In case of BM-25 and Doc2Vec, the difference in MAP between this two approaches is more than 0.1. This might be explained by the fact that the BM-25 model is not very suitable for very long queries. Other variants of BM-25 like BM-25L may need to be explored. Many of the documents contain more than 10-20 citations. In such a case, the total query text generated by the conjuction of sub-queries may significantly exceed query document size. When such a large amount of text is considered, the document may not represent all of its cited cases adequately. This may be one of the reasons for the drop in MAP in case of conjunction approach.

In case of LDA, we had already shown in our results, modeling an entire query document rather than a sub-query as a collection of topics gave better results. In concurrence with this observation, the Conjunction Approach which takes a single query with large amount of text has been able to generate fairly good topics as compared to max-pooling approach and thus giving slightly higher MAP value.

3 Runtime Performance of models for query retrieval

Here. all the models are taking citation context into consideration.

Table 2: Retrieval Time per query for all methods

Method Used	Runtime(sec)/query
BM-25	3.1
LDA	1.9
Doc2Vec	1.365
Word2Vec	0.686
TF-IDF	0.597

3.1 Analysis of results

From the results, we can see that BM-25 is the slowest among them. This is because BM-25 has to calculate scores for each query which is comparatively expensive process as compared to calculation of cosine similarity between pre-loaded prior case vectors and a query. All the cosine similarity based approaches such as TF-IDF, Word2Vec thus is comparatively faster. Even though Doc2Vec also involves cosine similarity, the generation of document embedding for each query document becomes a bottleneck, thus making it slower.

4 KL-Divergence vs Cosine Similarity for LDA

For topic distribution vectors obtained for LDA, a comparision of cosine similarity and KL-Divergence Score is performed.

Table 3: Comparing similarity metrics for topic distributions

Method Used	Metric	MAP	MRR	P@10	Rec@100
LDA	Cosine	0.103	0.205	0.745	0.508
LDA	KL Divergence	0.005	0.015	0.004	0.046

4.1 Analysis of results

From the results, we can see that that the cosine similarity approach has outperformed KL-divergence based metric significantly. The KL-divergence based approach didn't yield good results.