

Homework 4

1 Introduction

1.1 Collaboration and Originality

1. Did you receive help of any kind from anyone in developing your software for this assignment (Yes or No)? It is not necessary to describe discussions with the instructor or TAs.

No.

2. Did you give help of any kind to anyone in developing their software for this assignment (Yes or No)?

No.

3. Did you examine anyone else's software for this assignment (Yes or No)? It is not necessary to mention software provided by the instructor.

No.

4. Are you the author of every line of source code submitted for this assignment (Yes or No)? It is not necessary to mention software provided by the instructor.

Yes.

5. Are you the author of every word of your report (Yes or No)?

Yes.

1.2 Instructions

2 Experiment: Baselines

2.1 Experimental Results

	BM25 (Exp-1a)	Indri BOW (Exp-1b)	Indri SDM (Exp-1c)
P@10	0.2440	0.2440	0.2600
P@20	0.2620	0.2520	0.2480
P@30	0.2773	0.2840	0.2773
NDCG@10	0.1609	0.1538	0.1634
NDCG@20	0.1886	0.1708	0.1769
NDCG@30	0.2127	0.1983	0.2072
MAP	0.2138	0.1892	0.1926

2.2 Parameters

retrievalAlgorithm=Indri

BM25:k₁=1.2

BM25:b=0.75

BM25:k₃=0

Indri:mu=2500

Indri:lambda=0.4

For sequence dependency, I choose 0.8 for #AND operator, 0.1 for #NEAR operator and 0.1 for #WINDOW operator.

3 Custom Features

Feature 17, the date of the document is collected. This feature could be very helpful, because people might be more interested in new document than old document. The computation complexity is very low.

Feature 18 finds out all the positions of terms in the query and calculates its standard deviation. The hypothesis here is that if all terms appear together rather than diversely in the document, the document could be a better match. Because terms could be use together to form some concept appearing in the query. I use the standard deviation of all the positions to represent how closely terms are used in the document. The time complexity of creating this feature is in the order of the length of the document. Because a pass of scan of all positions is needed.

4 Experiment: Learning to Rank

4.1 Experiment Table

	BM25	IR Fusion (Exp-3a)	Content- Based (Exp-3b)	Base (Exp-3c)	All (Exp-3d)
P@10	0.2440	0.3920	0.4040	0.4120	0.3960
P@20	0.2620	0.3540	0.3540	0.3640	0.3660
P@30	0.2773	0.3400	0.3347	0.3493	0.2507
NDCG@10	0.1609	0.2586	0.2756	0.3039	0.2964
NDCG@20	0.1886	0.2644	0.2682	0.3104	0.3045
NDCG@30	0.2127	0.2838	0.2850	0.3162	0.3102
MAP	0.2138	0.2379	0.2394	0.2652	0.2581

4.2 Parameters

retrievalAlgorithm=LeToR

BM25:k₁=1.2

BM25:b=0.75

BM25:k₃=0

Indri:mu=2500

Indri:lambda=0.4

All Exp-3 experiments use the same queries to train the svm model but with different features.

4.3 Discussion

Comparing baseline and all other rerank output, we can see using svm to rerank the output has large improvement on MAP score, which shows that reranking is very useful. Comparing Exp-3d with Exp-3c, we get to know that not all features are useful in the reranking. And two customized features I create make the result worse.

Comparing Exp-3a to baseline, we know that combining BM25 and Indri scores together could help with the reranking. Comparing Exp-3b to Exp-3a, we know that using simple overlap feature could help with the reranking. Comparing Exp-3c to Exp-2b, document attributes like page rank and url depth are used and they have very good improvement.

5 Experiment: Feature Combinations

5.1 Experiment Table

	All (Baseline) (Exp-4a)	Comb₁ (Exp-4b)	Comb₂ (Exp-4c)	Comb₃ (Exp-4d)	Comb₄ (Exp-4e)
P@10	0.3960	0.3960	0.4320	0.4000	0.3960
P@20	0.3660	0.3760	0.3720	0.3760	0.3740
P@30	0.2507	0.3427	0.3560	0.3453	0.3320
NDCG@10	0.2964	0.2835	0.3191	0.2889	0.2888
NDCG@20	0.3045	0.3074	0.3163	0.3024	0.2992
NDCG@30	0.3102	0.3024	0.3248	0.3032	0.2908
MAP	0.2581	0.2574	0.2723	0.2523	0.2560

5.2 Parameters

retrievalAlgorithm=LeToR

BM25:k₁=1.2

BM25:b=0.75

BM25:k₃=0

Indri:mu=2500

Indri:lambda=0.4

Exp-4a: All features are used

Comb1: letor:featureDisable=10,13,16,17,18

Comb2: letor:featureDisable=14,15,16,17,18

Comb3: letor:featureDisable=7,10,11,12,14,15,16,17,18

Comb4: letor:featureDisable=1,3,14,15,16,17,18

5.3 Discussion

Comb1: According to Exp3c and Exp3d, I deactivate customized features 17 and 18. From feature 5 to feature 16, there are many scores calculated on same document field and query, whose effect could overlapping with each other. First, I try to remove the overlapScore on title, url and inlink. Because these scores already calculated by BM25 and Indri.

Comb2: This time I look into the effect of different fields. Inlink, comparing to other fields like body and title, could be useless in reranking. So, I remove the score for inlink to see what happens.

Comb3: Comb2 has good improvement. Based on Comb2, I keep removing features from 5 to 16. For body and title field. Using BM25 and Indri score could be better than simple OverlapScore, which is basically an Exact-Match score. So, I remove the OverlapScore for body and title field. For url field, because it is short, simple OverlapScore could be enough. So I remove BM25 and Indri score for url field.

Comb4: Based on previous best setting, Comb2, I try to test the usefulness of document attribute features. For feature1, document with high SpamScore might not matching the query user uses and spam websites could only be a small part in the corpus. So, this feature might not be so useful.

The final results show that SpamScore is useful in ranking, simple OverlapScore on body, title and url field is useful. Inlink field is not so useful in the ranking.

Coming to customized features, we can compare Exp3c and Exp3d. Both P@10 and MAP decreases after adding customized features.

6 Analysis

4a	4b	4c	4d	4e
1:1.2502978	1:1.2281497	1:1.2526499	1:1.2399714	1:
2:0.25389642	2:0.24225093	2:0.24536185	2:0.25426364	2:0.21998259
3:0.37919393	3:0.34878495	3:0.31685627	3:0.3232891	3:0.75461954
			4:-	
4:0.041397531	4:0.041441686	4:0.042200893	0.0029847675	4:0.29239419
5:1.2306018	5:1.2936056	5:1.2919636	5:1.4862512	5:1.445313
6:-0.061357122	6:-0.12387081	6:-0.064975105	6:-0.066056132	6:-0.78940713
7:0.63493705	7:0.60454738	7:0.64697361	7:	7:0.5648036
8:0.61380023	8:0.80882114	8:0.65891385	8:0.71423775	8:0.69303095
9:-0.31763595	9:-0.3068431	9:-0.29651827	9:-0.34418735	9:-0.25968707
10:0.17042969	10:	10:0.23518774	10:	10:0.35790306
11:-0.18876752	11:0.10785373	11:-0.18628779	11:	11:-0.23626238
12:-		12:-		12:-
0.095627882	12:0.063560717	0.088046819	12:	0.058867384
13:0.32639623	13:	13:0.33046147	13:0.20792772	13:0.32838061
14:-0.1503059	14:0.05901539			
15:0.034456141	15:0.028284648			
16:0.17656431				
17:-0.13290641				
18:-				
0.051822502				

From weight files output, we can see,

Useless features: 4(PageRank), 6(Indri score on body field), 12 (Indri score on url field), 14(BM25 score on inlink field), 15 (Indri score on inlink field)

Useful features: 1(SpamScore), 7 (Overlap score on body field), 8(BM25 score on body field), 9(Indri score on title field)

Exp4b confirms that Overlap score on body field is a useful feature. Exp4c confirms that BM25 and Indri score on inlink field is useless. Exp4d confirms that BM25 and Indir scores on url field is useful. Exp4e confirms that SpamScore is very useful.