Your Name Shuning Chen

Your Andrew ID shuningc

Homework 4

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. 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

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

Your Name Shuning Chen

Your Andrew ID shuningc

Homework 4

Instruction

1 Experiment: Baselines

	BM25	Indri BOW	Indri SDM
P@10	0.1680	0.1720	0.2080
P@20	0.2040	0.1980	0.2200
P@30	0.2053	0.1893	0.2293
MAP	0.0924	0.1146	0.1295

Baseline setting for BM25 is 0.2, 0.75,0 for k1,b and k3 respectively. Indri settings are 2500 and 0.4 for mu and lambda respectively.For indri SDM queries, I assigned weight 0.6,0.2 and 0.2 for and, near and window operators respectively.

2 Custom Features

Basic settings for BM25 and Indri are proved to perform effectively in previous assignments. For indri SDM queries, according to the experiments results in previous assignments, the higher the weight assigned to and operator, the better performance we get, it is more accurate to get exact match results. And for window and near operators, equal weight is assigned to help improve the total retrieval accuracy.

3 Experiment: Learning to Rank

	IR Fusion	Content-Based	Base	All
P@10	0.2680	0.2720	0.2000	0.2040
P@20	0.2440	0.2600	0.2180	0.2220
P@30	0.2387	0.2440	0.2213	0.2293
MAP	0.0978	0.0961	0.0844	0.0856

3.1 Parameters

Baseline setting for BM25 is 0.2, 0.75,0 for k1,b and k3 respectively. Indri settings are 2500 and 0.4 for mu and lambda respectively. For IR fusion,content-based, base and all models, the difference in setting is the number of disabled features. IR fusion disables 1,2,3,4,7,10,13,16,17,18. Based model disables 1,2,3,4,17,18. Base model disables 17 and 18.

3.2 Discussion

Comparing the performance produced by IR fusion and content-based model, it seems like overlap score doesn't improve the overall performance. On the contrary, it makes the performance worse. It is not the same as I expected. I expect that the more features, the better the performance. It means the BM25 and Indri are more accurate than the other features such as overlap score. Only the first model overperforms baseline models. It might be the situation that BM25 and Indri both pay more attention to exact match, and so it gains higher retrieval result for precision. Even within the training models, feature BM25 and Indri score gives better feedback. In terms of computational complexity, BM25 and Indri are more complicated to compute compared to other features. From this prospective, the proved complicated computation produces rather good predictions.

4 Experiment: Features

Experiment with four different combinations of features.

	All (Baseline)	**5,8,11,14**	**5,6,8,9,11, 12,14,15**	**6,9,12,15**	**5,6,7,8,9,10 ,11,12,13,1 4,15,16**
P@10	0.2160	0.2120	0.2680	0.2840	0.2000
P@20	0.2220	0.2460	0.2440	0.2660	0.2180
P@30	0.2347	0.2440	0.2387	0.2560	0.2213
MAP	0.0848	0.0926	0.0978	0.0963	0.0844

4.1 Parameters

Baseline setting for BM25 is 0.2, 0.75,0 for k1,b and k3 respectively. Indri settings are 2500 and 0.4 for mu and lambda respectively.

Comb1 enables 5,8,11,14. These are BM25 scores. Comb2 enables 5,6,8,9,11,12,14,15. These are BM25 and Indri scores. Comb3 enables 6,9,12,15, and these are Indri scores. Comb4 enables 5,6,7,8,9,10,11,12,13,14,15,16. These are BM25, Indri and overlap scores.

4.2 Discussion

Comb1 enables BM25 scores. Comb2 enables BM25 and Indri scores. Comb3 enables Indri scores. Comb4 enables BM25, Indri and overlap scores. Comb3 has the highest computational complexity because it calculated scores for both BM25 and Indri which are the most complicated to calculate among all. I chose the combination of BM25, Indri, Indri and BM25, Indri, BM25 and overlap score respectively.

According to the results of previous experiment results, BM25 and Indri tend to give better performance. So I tried to train the model with BM25 scores only and with Indri scores only. And it seems the combination of feature BM25 scores and Indri scores works better. Compared to combinations of all features and based features, it has a better score with smaller set of features containing only BM25 and Indri features.Compared to BM25 features only or Indri features only, a larger set with combination of them works better.

BM25 and Indri all give the score for exact match, it is the mathematical calculation with some smoothing makes prediction more accurate. However, other features such as pagerank and spam scores only give a percentage score, which might not give the training model a good indication.

5 Analysis

Observed from the models produced, it shows that the spam score and wikipedia scores takes a lot of account. It appears that these two are more useful than others. But this cannot be reflected in my experiments. BM25 and Indri scores are all very important, but it appears that BM25 is more useful according to the scores generated. These can also be observed from my experiments' results. The difference is that my experiments didn't pay attention to the difference of importance between BM25 and Indri. And for overlap score in the inlink field, the model shows that it can make a difference. It doesn't represent that overlap score for all fields are useful. Overall, weight assigned to BM25 and Indri is consistent with my experiments' results.