Query Performance Prediction

Your name

ABSTRACT

Only template

Write your abstract here. Your abstract should *concisely* say (i) *why* the topic is interesting, (ii) *what* you do in your study, (iii) *how* you did your study and (iv) *what* the results were

1. INTRODUCTION

The world is heavily dependant on search engines. Because of this, it is of increasing importance that search engines provide the best possible result for all queries. This leads to increasing demands for information retrieval (IR) systems, when predicting the succesfullnes of new approaches.

Query performance prediction (QPP) is one approach for testing approaches. This approach is used to calculate the performance of a query before or after it's put through the IR system. Using QPP it is possible to choose strategy with a queries to improve the IR systems performance and succesfullnes. One way would be to preemptivly try to predict how a query would perform with several different search engines. Then the system could choose the one ir predict it would perform the best width.

There are several studies about QPP, and a study by Hauff and Azzopardi[**?**] test whether or not there is actually any need for us to apply QPP. They investigate when there is a reason to even try to apply QPP on search queries to increse performance. They conclude that even QPP can increase IR systems performance, it only applies when Kendall's Tau correlation of $\tau >= 0.5$

Only template

Write your introduction here. Get inspiration on how to structure and formulate an introduction from the studies you review. Make sure you describe what query performance prediction is and why it is useful.

Several studies on query-performance prediction have been published before and after [1]. In your introduction, give an overview of ***at least three*** studies of query-performance prediction published by the ACM between 2005 and 2016. You may find these studies by searching e.g. the ACM digital library (http://dl.acm.org) or Google Scholar. Your literature review of the 3+ papers must (i) describe the method proposed in the paper, how the method was evaluated and what the results were. Furthermore, it should be clear how each paper differ from the other paper you review. **Remember: the literature review is meant to help the reader understand where there is a gap in the existing research that you can fill**. Therefore, select papers that are as close as possible to [1].

You *must* cite your sources when/if you use a specific phrasing. Failure to do so will be considered plagiarism. aaaaa

2. DATA SETS & QUERIES

2.1 Test collections

The data used in this paper is a collection of TREC collections with a large amount of documents, and varying document length and length deviation to test the prediction in many situations. Table 1 shows an overview of the data.

Table 1: Document data sets

Name	laTimes	fbis	ft
document count	131896	130471	210158
average length	502	504	399
minimum length	2	12	11
maximum	24125	143175	26145

The queries look like this

Table 2: queries data set

Name	queries
document count	150
average length	19.54

Furthermore, list the number of queries and the average query length for the superset of the queries.

3. EXPERIMENTS

Only template

For indexing, retrieval and evaluation use INDRI and use TREC_EVAL. Index the data sets using the Krovetz stemmer and stop word removal using the list http://ir.dcs.gla.ac.uk/resources/linguistic_utils/stop_words. Use the

Information Retrieval 2016 DIKU, Denmark
.

BM25 ranking model for retrieval. You are not required to tune any parameter, but are welcome to do so.

Use values of $N = \{20, 50, 100, 200\}$ and report a table similar to that of Table 2 in [1] setting $\sigma = \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and include the normalised version (see the paper). Use both Pearson's r and Spearman's ρ, and use only the title of each query. Where is the correlation with P@10 the highest for each N? Where is it best overall? Why do you think that is? Repeat the above with at least one other metric (e.g. MRR, nDCG etc.). Which one gives you the highest correlation?

As the standard deviation assumes data are normally distributed (which is not case for many real-life data sets), repeat the above analysis, but instead of σ use the *mean absolute deviation*:

$$\mathrm{MAD} = \mathrm{median}\left(x_i - \mathrm{median}(x)\right) \tag{1}$$

Does MAD correlate with P@10 better than compared to using the standard deviation? Why? What is the best correlation you get? What about the normalised version of the MAD?

Finally, produce a Table similar to Table 4 in [1] for one of the datasets of your choosing. Use the simplified clarify score (scs), average IDF (idf_{avg}), query scope (qs) and $\sigma_{\max}$ predictors.

The simplified clarify score [?] (scs) is given by:

$$scs(Q) = \sum_{w \in V} P(w|Q) \log_2 \frac{P(w|Q)}{P_{\mathrm{coll}(w)}} \tag{2}$$

where w is a word in the vocabulary V of the data set or query Q, $P(w|Q)$ is the conditional probability (relative frequency or MLE estimate) of w in the query Q and $P_{\mathrm{coll}}(w)$ is the probability of w in the data set (the relative frequency of w in V).

The query scope [?] (qs) is given by:

$$qs(Q) = -\log(n_Q/N) \tag{3}$$

where log denotes the natural logarithm, n_q is the number of documents in the data set that contains *at least* one of the query terms, and N is the number of documents in the data set.

Compare your best correlations with these predictors.

4. TABLES

Table 3: $\sigma_{\%}$ correlations for la20

la 20	$\sigma_{0.1\%}$	$\sigma_{0.3\%}$	$\sigma_{0.5\%}$	$\sigma_{0.7\%}$	$\sigma_{0.9\%}$
Pearsons	0.3686	0.3734	0.3665	0.2312	0.1115
Spearman	0.3464	0.3464	0.3625	0.2578	0.0409

Table 4: $\sigma_{\%}$ correlations for la50

la 50	$\sigma_{0.1\%}$	$\sigma_{0.3\%}$	$\sigma_{0.5\%}$	$\sigma_{0.7\%}$	$\sigma_{0.9\%}$
Pearsons	0.3824	0.3698	0.4077	0.2335	0.1067
Spearman	0.2571	0.253	0.3512	0.2385	0.0393

Table 5: $\sigma_{\%}$ correlations for la100

la 100	$\sigma_{0.1\%}$	$\sigma_{0.3\%}$	$\sigma_{0.5\%}$	$\sigma_{0.7\%}$	$\sigma_{0.9\%}$
Pearsons	0.2708	0.3171	0.3954	0.2268	0.1042
Spearman	0.2192	0.2798	0.3305	0.2319	0.0368

Table 6: $\sigma_{\%}$ correlations for la200

la 200	$\sigma_{0.1\%}$	$\sigma_{0.3\%}$	$\sigma_{0.5\%}$	$\sigma_{0.7\%}$	$\sigma_{0.9\%}$
Pearsons	0.2762	0.3543	0.3608	0.2172	0.1042
Spearman	0.1418	0.2061	0.2512	0.2182	0.0368

Table 7: $\sigma_{\%}$ correlations for ft20

ft 20	$\sigma_{0.1\%}$	$\sigma_{0.3\%}$	$\sigma_{0.5\%}$	$\sigma_{0.7\%}$	$\sigma_{0.9\%}$
Pearsons	0.2388	0.2277	0.2185	0.277	0.2452
Spearman	0.3025	0.3032	0.3437	0.32	0.1657

Table 8: $\sigma_{\%}$ correlations for ft50

ft 50	$\sigma_{0.1\%}$	$\sigma_{0.3\%}$	$\sigma_{0.5\%}$	$\sigma_{0.7\%}$	$\sigma_{0.9\%}$
Pearsons	0.1693	0.1787	0.2528	0.2633	0.2411
Spearman	0.2317	0.2458	0.3577	0.2708	0.1622

Table 9: $\sigma_{\%}$ correlations for ft100

ft 100	$\sigma_{0.1\%}$	$\sigma_{0.3\%}$	$\sigma_{0.5\%}$	$\sigma_{0.7\%}$	$\sigma_{0.9\%}$
Pearsons	0.0658	0.1572	0.2635	0.2543	0.2395
Spearman	0.1196	0.1869	0.3527	0.2418	0.1585

Table 10: $\sigma_{\%}$ correlations for ft200

ft 200	$\sigma_{0.1\%}$	$\sigma_{0.3\%}$	$\sigma_{0.5\%}$	$\sigma_{0.7\%}$	$\sigma_{0.9\%}$
Pearsons	-0.0021	0.1208	0.2786	0.2491	0.2395
Spearman	0.0238	0.1237	0.3543	0.2232	0.1585

Table 11: $\sigma_{\%}$ correlations for fbis20

fbis 20	$\sigma_{0.1\%}$	$\sigma_{0.3\%}$	$\sigma_{0.5\%}$	$\sigma_{0.7\%}$	$\sigma_{0.9\%}$
Pearsons	0.2852	0.2735	0.2008	0.105	0.3691
Spearman	0.1882	0.1807	0.1583	0.111	0.335

Table 12: $\sigma_{\%}$ correlations for fbis50

fbis 50	$\sigma_{0.1\%}$	$\sigma_{0.3\%}$	$\sigma_{0.5\%}$	$\sigma_{0.7\%}$	$\sigma_{0.9\%}$
Pearsons	0.1314	0.1447	0.1763	0.1028	0.3706
Spearman	0.1871	0.1692	0.2139	0.0994	0.3439

Table 13: $\sigma_{\%}$ correlations for fbis100

fbis 100	$\sigma_{0.1\%}$	$\sigma_{0.3\%}$	$\sigma_{0.5\%}$	$\sigma_{0.7\%}$	$\sigma_{0.9\%}$
Pearsons	0.2276	0.28	0.2434	0.1127	0.3706
Spearman	0.2127	0.2153	0.2495	0.1019	0.3456

Table 14: $\sigma_{\%}$ correlations for fbis200

fbis 200	$\sigma_{0.1\%}$	$\sigma_{0.3\%}$	$\sigma_{0.5\%}$	$\sigma_{0.7\%}$	$\sigma_{0.9\%}$
Pearsons	0.2032	0.2921	0.2416	0.1049	0.3704
Spearman	0.232	0.2715	0.2545	0.1063	0.3452

5. CONCLUSION

Write your conclusion here

6. REFERENCES

[1] R. Cummins, J. Jose, and C. O'Riordan. Improved

query performance prediction using standard deviation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 1089–1090. ACM, 2011.