Exploiting Query Reformulations for Web Search Result Diversification

Rodrygo L. T. Santos
Department of Computer Science
University of Glasgow, UK

Craig Macdonald
Department of Computer Science
University of Glasgow, UK

Iadh Ounis
Department of Computer Science
University of Glasgow, UK

Presented By

Wasi Uddin Ahmad

Md Masudur Rahman

13th April, 2016

UNIVERSITY of VIRGINIA

Motivation

- Java
  - ‘java programming language’
  - ‘java’ – an island of Indonesia
  - ‘java coffee’
- What if an ambiguous query is submitted to the search engine?
  - Completely ignore any sort of ambiguity
  - Infer the most plausible meaning underlying the query
  - Explicitly ask the user for feedback on the correct meaning underlying the query
  - Diversify the retrieved results of the query

Diversifying Search Result

- Given an initial ranking R for a query q, find a re-ranking S that has the *maximum coverage* and the *minimum redundancy* with respect to the different aspects underlying q
- How to diversify search results?
  - Compare the retrieved documents for a given query to one another
  - Select the documents most relevant to the query while being the most dissimilar to the documents already selected
  - Assumption – similar documents will cover similar aspects underlying the query and should be demoted in order to achieve diversified ranking

Related Work

- Implicit approaches
  - Similar documents will cover similar aspects and should hence be demoted
- Explicit approaches
  - Directly models the query aspects
  - Maximize the coverage of the selected documents with respect to these aspects

Implicit Approaches

- Carbonell and Goldstein [MMR] – selects document based on the combination of a similarity and a dissimilarity score
  - Content based similarity function
- Zhai and Lafferty – used language modeling framework
- Chen and Karger – proposed a probabilistic approach
- Wang and Zhu – employed correlation between documents as a measure of similarity

Explicit Approaches

- Agarwal et al. [IA Select] used a taxonomy for both queries and documents
  - Two documents are similar if they are classified into one or more common categories covered by the query
- Carterette and Chandar – proposed a probabilistic model
  - To maximize the coverage of a document ranking with respect to query aspects
- Radlinski and Dumais [Q-Filter] – proposed to filter the document ranking
  - To have a more even distribution of documents satisfying each query aspect

Contribution of the paper

- Follows the explicit approach
- Novel probabilistic framework for search result diversification
  - models the information need of an ambiguous query as a set of sub-queries
- Analysis of the effectiveness of the sub-queries
  - Derived from two types of query reformulation provided by three major WSE
- Thorough evaluation of the several components of the proposed framework

Main Framework

xQuAD(q, R, τ, λ)

$S \leftarrow \emptyset$

while $|S| < \tau$ **do**

  $d^* \leftarrow \arg\max_{d \in R \setminus S} \boxed{(1-\lambda)\,\mathrm{P}(d|q) + \lambda\,\mathrm{P}(d, \bar{S}|q)}$

  $R \leftarrow R \setminus \{d^*\}$

  $S \leftarrow S \cup \{d^*\}$

end while

return S

Algorithm 1: The xQuAD framework.

UNIVERSITY of VIRGINIA

xQuAD Framework

$$(1-\lambda)\boxed{\mathrm{P}(d|q)}+\lambda\boxed{\mathrm{P}(d,\bar{S}|q)},$$

Document query relevance

Maximum coverage
Minimum redundancy

- q = ambiguous query
- R = initial ranking produced for query, q
- S = new ranking by iteratively selecting highest scored documents from R
- $P(d|q)$ = likelihood of document d being observed given q
- $P(d,\bar{S}|\, q)$ = likelihood of observing this document but not the document already in S

xQuAD Framework

$$\mathrm{P}(d, \bar{S}|q) = \sum_{q_i \in Q} \mathrm{P}(q_i|q)\, \mathrm{P}(d, \bar{S}|q_i),$$

- $P(q_i|q)$ = measure of the relative importance of the sub-query q_i

$$\mathrm{P}(d, \bar{S}|q_i) = \mathrm{P}(d|q_i)\, \mathrm{P}(\bar{S}|q_i),$$

- $P(d|q_i)$ = measure of the coverage of document d with respect to the sub-query q_i
- $P(\bar{S}|q_i)$ = measure of novelty; the probability of q_i not being satisfied by any of the documents already selected in S

xQuAD Framework

$$\mathrm{P}(\bar{S}|q_i) = \mathrm{P}(\overline{d_1, \cdots, d_{n-1}}|q_i)$$
$$= \prod_{d_j \in S} (1 - \mathrm{P}(d_j|q_i)).$$

- Assumption
  - Relevance of a document in S to a given sub-query q_i is independent of the relevance of other documents in S to the same sub-query
- Final Equation becomes,

$$(1-\lambda)\,\mathrm{P}(d|q) + \lambda \sum_{q_i \in Q} \left[\mathrm{P}(q_i|q)\,\mathrm{P}(d|q_i) \prod_{d_j \in S} (1 - \mathrm{P}(d_j|q_i))\right].$$

Components Estimation

- Document relevance, Coverage and Novelty
  - Any probabilistic approach can be used, e.g., language modeling
  - Document ranking for the initial query [baseline ranking]
  - Ranking produced for the sub-queries [sub-rankings]
- Sub-Query Generation
  - Traditional query expansion techniques in order to generate ‘expanded sub-queries’
  - Using search query log, possible search queries can be generated
  - Using *related sub-queries* and *suggested sub-queries*

Components Estimation

- Sub-Query Importance, $P(q_i|q)$
  - Baseline estimation – all sub-queries are equally important $\mathrm{P}_u(q_i|q) = \frac{1}{|Q|},$
  - Relative importance of each sub-query based on how well it is covered by a given collection

$$\mathrm{P}_w(q_i|q) = \frac{n_w(q_i)}{\sum_{q_j \in Q} n_w(q_j)},$$

  - CRCS based sub-query importance estimation

$$i_c(q_i|q) = \frac{n_c(q_i)}{\max_{q_i \in Q} n_c(q_i)} \frac{1}{\hat{n}_c(q_i)} \sum_{d|\,\mathrm{P}(d|q_i)>0} \tau - j(d,q),$$

$$\mathrm{P}_c(q_i|q) = \frac{i_c(q_i|q)}{\sum_{q_j \in Q} i_c(q_j|q)}.$$

Experimental Setup

- Collection and Topics
  - A subset of TREC ClueWeb09 dataset was used
  - 50 topics were used where each topic includes 3 to 8 sub-topics
- Evaluation Metrics
  - α-NDCG and IA-P (intent-aware precision)
  - Three different rank cutoffs: 5, 10, and 100
- Retrieval Baselines
  - BM25, DPH and LM (language modeling)
- Training Procedures
  - In order to train λ, 5-fold cross validation over the 50 topics was performed

Experimental Evaluation

	α-NDCG			IA-P		
	@5	@10	@100	@5	@10	@100
BM25	0.159	0.186	0.288	0.075	0.071	**0.059**
+MMR	0.120	0.150	0.224	0.056	0.058	0.039
+Q-Filter	0.159	0.186	0.286	0.075	0.071	0.057
+IA-Select	0.110	0.119	0.180	0.043	0.037	0.023
+xQuAD$_u$	**0.208**	**0.227**	**0.324**	**0.080**	**0.075**	0.056
DPH	0.198	0.212	0.304	**0.109**	**0.106**	**0.062**
+MMR	0.195	0.211	0.303	0.105	0.103	**0.062**
+Q-Filter	0.198	0.212	0.303	**0.109**	**0.106**	0.060
+IA-Select	0.148	0.157	0.203	0.077	0.071	0.023
+xQuAD$_u$	**0.208**	**0.243**	**0.334**	0.097	0.096	0.061
LM	0.082	0.096	0.180	0.041	0.040	0.032
+MMR	0.083	0.096	0.183	0.041	0.039	0.032
+Q-Filter	0.078	0.095	0.179	0.040	0.040	0.031
+IA-Select	0.081	0.086	0.127	0.037	0.027	0.014
+xQuAD$_u$	**0.085**	**0.104**	**0.198**	**0.045**	**0.042**	**0.034**

Table 2: Diversification performance using the official TREC 2009 Web track diversity sub-topics.

Experimental Evaluation

		related sub-queries						suggested sub-queries					
		α-NDCG			IA-P			α-NDCG			IA-P		
	WSE	@5	@10	@100	@5	@10	@100	@5	@10	@100	@5	@10	@100
BM25		0.159	**0.186**	**0.288**	0.075	0.071	**0.059**	0.159	**0.186**	0.288	0.075	**0.071**	**0.059**
+xQuAD$_u$	A	0.154	0.184	0.282	0.070	0.072	0.057	**0.171**	**0.186**	**0.291**	0.082	**0.071**	0.053
+xQuAD$_u$	B	0.154	0.182	0.279	0.073	**0.076**	0.054	0.129	0.158	0.261	0.065	0.067	0.052
+xQuAD$_u$	C	**0.161**	0.182	0.285	**0.076**	**0.076**	0.057	0.163	0.184	0.287	**0.084**	0.069	0.053
DPH		0.198	**0.212**	0.304	**0.109**	**0.106**	**0.062**	0.198	0.212	0.304	**0.109**	**0.106**	**0.062**
+xQuAD$_u$	A	0.164	0.189	0.288	0.086	0.083	0.056	**0.215**	0.222	0.313	0.108	0.088	0.055
+xQuAD$_u$	B	0.186	0.205	0.295	0.090	0.082	0.057	0.162	0.189	0.281	0.088	0.085	0.055
+xQuAD$_u$	C	**0.206**	0.209	**0.307**	0.108	0.090	**0.062**	0.201	**0.236**	**0.320**	0.093	0.092	0.059
LM		0.082	0.096	0.180	**0.041**	0.040	0.032	0.082	0.096	0.180	0.041	0.040	0.032
+xQuAD$_u$	A	**0.088**	0.103	**0.192**	0.038	0.038	0.032	**0.101**	0.123	0.204	0.043	0.046	0.032
+xQuAD$_u$	B	0.081	**0.105**	0.188	0.040	**0.045**	**0.033**	0.093	0.118	0.197	0.041	0.043	0.033
+xQuAD$_u$	C	0.082	0.100	0.183	0.037	0.039	0.032	**0.101**	**0.127**	**0.205**	**0.046**	**0.047**	**0.034**

Table 3: Diversification performance using related and suggested sub-queries from different WSEs.

Experimental Evaluation

	α-NDCG			IA-P		
	@5	@10	@100	@5	@10	@100
BM25	0.159	0.186	0.288	0.075	0.071	**0.059**
+xQuAD$_u$	**0.208**	**0.227**	**0.324**	**0.080**	**0.075**	0.056
+xQuAD$_c$	0.176	0.206	0.296	0.066	0.066	0.048
+xQuAD$_w$	0.184	0.201	0.297	0.077	0.067	0.053
DPH	0.198	0.212	0.304	**0.109**	**0.106**	**0.062**
+xQuAD$_u$	**0.208**	**0.243**	**0.334**	0.097	0.096	0.061
+xQuAD$_c$	0.169	0.204	0.299	0.073	0.073	0.053
+xQuAD$_w$	0.203	0.226	0.316	0.101	0.088	0.060
LM	0.082	0.096	0.180	0.041	0.040	0.032
+xQuAD$_u$	0.085	0.104	0.198	**0.045**	0.042	0.034
+xQuAD$_c$	**0.110**	**0.146**	**0.234**	0.044	**0.047**	**0.041**
+xQuAD$_w$	0.078	0.095	0.187	0.039	0.039	0.033

Table 4: Diversification performance using different sub-query importance estimators.

Conclusion and Future Works

- A novel probabilistic framework for search result diversification
- Thoroughly experimented the effectiveness of the framework
- Future works
  - More effective sub-query generation
  - More sophisticated document retrieval techniques might improve relevance, coverage and novelty components

Any Question?

Thank You

University of Virginia