Relevance and Diversity for Retrieval

Kuan-Yu Chen
Institute of Information Science, Academia Sinica
Reference


Outline

• Introduction

• Various Diversification Methods
  – Maximal Marginal Relevance (MMR)
  – Simple Mixture Model (SMM)
  – WUME
  – eXplicit Query Aspect Diversification (xQuAD)

• Analytical Comparisons

• Experimental Results
Introduction — What’s going on?

- Traditional retrieval functions ignore the relations among returned documents
  - Top ranked documents may contain relevant yet redundant information
  - In order to maximize the satisfaction of different search users, it is necessary to diversify search results
Introduction — Various Modeling

• Many diversification methods have been proposed
  – balance the relevance and the redundancy: MMR
  – distinguish previous topics and new coming: SMM
  – language modeling approach: WUME
  – probabilistic framework: xQuAD

• These methods mainly differ in diversity modeling
  – Implicitly: The diversity is implicitly modeled through document similarities
  – Explicitly: It can be explicitly modeled through the coverage of query subtopics, and document dependency
## Introduction — Notations

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Q )</td>
<td>A given query</td>
</tr>
<tr>
<td>( q_i )</td>
<td>Sub-queries, ( Q = {q_1, \ldots, q_K} )</td>
</tr>
<tr>
<td>( K )</td>
<td>Number of sub-queries</td>
</tr>
<tr>
<td>( R )</td>
<td>The user’s information need</td>
</tr>
<tr>
<td>( D )</td>
<td>A set of documents, ( D = {d_1, \ldots, d_M} )</td>
</tr>
<tr>
<td>( M )</td>
<td>Number of documents</td>
</tr>
<tr>
<td>( \tilde{D} )</td>
<td>A subset of documents which already selected by new method, ( \tilde{D} = {\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_N} )</td>
</tr>
<tr>
<td>( N )</td>
<td>Number of selected documents</td>
</tr>
</tbody>
</table>
Introduction — General Framework

• All these methods iteratively select the document that is not only relevant to the query but also diversified to cover more query subtopics, explicitly or implicitly.

• All of methods fit into a general framework that iteratively selects with the highest relevance and diversity scores:

\[ d^* = \arg\max_{d \in D} \lambda \cdot \text{Rel}(d, Q) + (1 - \lambda) \cdot \text{Div}(d, Q) \]
MMR

• MMR motivated the need for “relevant novelty” as a potentially superior criterion
  – An approximation to measuring relevant novelty is to measure relevance and novelty independently

• “Marginal Relevance” can be regarded as the metric
  – A document has high marginal relevance if it is both relevant to the query and contains minimal similarity to previously selected documents

\[ \text{Div}_{\text{MMR}}(d, Q) = -\max_{\tilde{d} \in \tilde{D}} \text{Sim}(\tilde{d}, d) \]
SMM

• Given the observed new document, we estimate the mixing weight for the background model $\theta_B$ and the previous topic model $\theta_T$
  
  – The mixture weight for the background model can serve as a measure of novelty or redundancy

  $$L(\lambda | d, \theta_B, \theta_T) = \sum_{w_i \in V} [(1 - \lambda)P(w_i | \theta_T) + \lambda P(w_i | \theta_B)]c(w_i, d)$$

  $$\text{Div}_{\text{SMM}}(d, Q) = \arg\max_{\lambda} L(\lambda | d, \theta_B, \theta_T)$$

• The simplest previous topic model can be modeled as:

  $$P(w_i | \theta_T) = \sum_{\tilde{d}_n \in \tilde{D}} \frac{1}{N} P(w_i | \tilde{d}_n)$$
WUME — The basic motivation

- There are three sub-queries under the given query $q_1$, $q_2$, and $q_3$, and web documents $d_1$ through $d_8$.

- Although $d_3$ is more relevant to one of the sub-query $q_2$ than $d_5$ to $q_3$, given that $q_2$ attracts less user interest than $q_3$, $d_3$ should still be ranked lower than $d_5$. 
WUME

• WUME formalize the diversification method as:
  – Given a query $Q$, the probability that a retrieved document meets user’s information need $R$ can be written as:

    $$P(R \mid Q,d) = \frac{P(R \mid Q)P(d \mid R,Q)}{P(d \mid Q)}$$

  – WUME assume that, without considering user’s information need $R$, the probability of each retrieved document given the query is the same across all retrieved documents

    $$P(R \mid Q,d) \propto P(d \mid R,Q)$$
WUME

– Now we take sub-query information into consideration, where \( q_t \) represents a sub-query associated with query \( Q \)

\[
P(d \mid R, Q) = P(d \mid q_1, R, Q) \times P(q_1 \mid R, Q)
+ \cdots
+ P(d \mid q_k, R, Q) \times P(q_k \mid R, Q)
= \sum_{t=1}^{k} P(d \mid q_t, R, Q) \times P(q_t \mid R, Q)
\]

– Finally, WUME have

\[
\text{Div}_{\text{WUME}}(d, Q) = \sum_{t=1}^{k} P(d \mid q_t, R, Q) \times P(q_t \mid R, Q)
\]
xQuAD

• When given an ambiguous query, xQuAD builds a new ranked list by:

\[ d^* = \arg\max_d (1 - \lambda)P(d \mid Q) + \lambda P\left(d, \tilde{D} \mid Q\right) \]

  – \( P(d \mid Q) \) is the likelihood of document \( d \) being observed given the initial query
    • The probability can be regarded as modeling \textit{relevance}

  – \( P\left(d, \tilde{D} \mid Q\right) \) is the likelihood of observing this document but not the documents already in \( \tilde{D} \)
    • The probability can be regarded as modeling \textit{diversity}
xQuAD

• In order to derive $P(d, \tilde{D} | Q)$, xQuAD explicitly consider the possibly several aspects underlying the initial query as a set of sub-queries.

• By enforcing $\sum_{q_t \in Q} P(q_t | Q) = 1$, xQuAD can marginalize $P(d, \tilde{D} | Q)$ across multiple sub-queries:

$$P(d, \tilde{D} | Q) = \sum_{q_t \in Q} P(d, \tilde{D} | q_t) P(q_t | Q)$$

• Next, $P(d, \tilde{D} | q_t)$ can be broken down by independent assumption:

$$P(d, \tilde{D} | q_t) = P(d | q_t) P(\tilde{D} | q_t)$$
xQuAD

• The independence assumption has a subtle but important implication
  – It turns the computation of novelty from a direct comparison between documents into an estimation of the marginal utility of the sub-queries satisfied by a document
    – In other words, instead of comparing a document $d$ to all documents already selected in $\tilde{D}$, as implicit diversification approaches would do, xQuAD estimates the utility of any document satisfying the sub-query $q_i$, given how well it is already satisfied by the documents in $\tilde{D}$
xQuAD

- xQuAD also assumes that the relevance of each document in $\tilde{D}$ to a given sub-query $q_i$ is independent

$$P\left(\tilde{D} \mid q_t\right) = P\left(\tilde{d}_1, \ldots, \tilde{d}_N \mid q_t\right)$$

$$= \prod_{\tilde{d}_j \in \tilde{D}} \left(1 - P\left(\tilde{d}_j \mid q_t\right)\right)$$

- To sum up, xQuAD suggests that:

$$\text{Div}_{xQuAD}(d, Q) = \sum_{t=1}^{K} P(q_t \mid Q) P(d \mid q_t) \prod_{\tilde{d}_j \in \tilde{D}} \left(1 - P\left(\tilde{d}_j \mid q_t\right)\right)$$
Analytical Comparisons

• Diversity Modeling:
  – MMR and SMM implicitly models the diversity through document similarities and ignores the information about query subtopics
  – The other two methods explicitly model the diversity through the coverage of query subtopics

• Document Dependency:
  – WUME assumes that the diversity score of a document is independent of other documents
  – The other three methods assume that the diversity score depends on the previously selected documents
Experimental Results

- All the parameters in each method are set to the optimum values
  - Both xQuAD and WUME perform significantly better than MMR
    - Using explicit sub-queries in diversification is better than implicit sub-queries
  - The performances of xQuAD and WUME are not significantly different
    - The component of sub-queries importance penalization of xQuAD needs to be modified

<table>
<thead>
<tr>
<th></th>
<th>TREC09 result</th>
<th>TREC10 result</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α-nDCG@20</td>
<td>α-nDCG@100</td>
</tr>
<tr>
<td>MMR*</td>
<td>0.365</td>
<td>0.427</td>
</tr>
<tr>
<td>WUME*</td>
<td>0.479</td>
<td>0.546</td>
</tr>
<tr>
<td>xQuAD*</td>
<td>0.482</td>
<td>0.550</td>
</tr>
</tbody>
</table>
Conclusions

• The experiment result shows that the explicit sub-query modeling and sub-query importance penalization strategies perform better

• It is interesting to find that how the sub-queries affect the overall performance

• Finally, we can think about that what’s difference between sub-queries and latent topics?

• Beyond relevance or another relevance?