UMass Robust 2005 Notebook:
Using Mixtures of Relevance Models for Query Expansion

Donald Metzler, Fernando Diaz, Trevor Strohman, W. B. Croft
Center for Intelligent Information Retrieval
University of Massachusetts, Amherst

Abstract
This paper describes the UMass TREC 2005 Robust
Track experiments. We focus on approaches that use
term proximity and pseudo-relevance feedback using
external collections. Our results indicate both ap-
proaches are highly effective.

1 Introduction
For the 2005 Robust Track, we explore whether or not
term proximity information and advanced pseudo-
relevance feedback methods can be used to achieve
good effectiveness on a challenging query set.

All experiments used the Indri search engine [3],
indexed the full AQUAINT collection of 1,033,461
documents, used a Porter Stemmer and a stopword
list of 418 common terms. All runs are automatic.

2 Dependence Model
We use Metzler’s dependence model formulation
to exploit term proximity information, which has been
shown to significantly improve effectiveness over sim-
ple bag of words models [2]. The Indri query language
can be used to express dependence model queries.
This helps give an intuitive meaning to the model.

For example, for topic 625, “arrests bombing wtc”,
the following Indri query ranks documents exactly as
done by the dependence model:

\[ P(w|Q) = \sum_{c \in C} P(c|Q)P(w|Q,c) \]
\[ \propto \sum_{c \in C} P(c|Q) \int_{\theta} P(w|\theta)P(Q|\theta)P(\theta|c) \]

In order to make evaluation of this expression more
feasible, we follow Lavrenko [1] and approximate the
integral by a summation over the models of the top
ranked documents. We denote these models as \( \mathcal{R}_c \),

\[ #weight(0.8 #combine(arrests bombing wtc) \]
\[ 0.1 #combine(#1(arrests bombing) \]
\[ #1(bombing wtc) \]
\[ #1(arrests bombing wtc)) \]
\[ 0.1 #combine(#uw8(arrests bombing) \]
\[ #uw8(bombing wtc) \]
\[ #uw6(bombing wtc) \]
\[ #uw12(arrests bombing wtc)) \]

From this formulation we see that proximity infor-
mation, in the form of exact phrases (#1) and un-
ordered windows (#uwN) play a vital role in how doc-
uments are ranked.

3 Mixture of Relevance Models
Lavrenko’s relevance models are a powerful way to
construct a query model from a set of top ranked
documents [1]. We generalize the idea to allow evi-
dence to be incorporated from multiple collections.
We take a Bayesian approach, and see that:
where the subscript indicates the collection. Furthermore, we also assume that \( P(\theta|c) = \frac{1}{|K_c|} \) and that
\( P(c|Q) = P(c) \) for all \( Q \), which implies the mixture weights are equal for every query. Better distributional assumptions for \( P(\theta|c) \) and actually computing \( P(c|Q) \) may lead to better estimates, but is left as future work. Under these simplifying assumptions, we get the following estimate for our query model:

\[
P(w|Q) \propto \sum_{c \in C} \frac{P(c)}{|R_c|} \sum_{\theta \in \Theta_c} P(w|\theta)P(Q|\theta)
\]

where we tune \(|R_c| \) and \( P(c) \) on training data.

Now that we have a query model that combines evidence from multiple collections, we can use it for query expansion by adding the \( k \) most likely terms from the distribution \( P(w|Q) \) to the original query.

In our experiments, we investigate mixing models from two collections, AQUAINT, and BIGNESWS, a collection of 6,160,058 TREC newswire articles we had on site.

4 Effectiveness Prediction

For predicting query effectiveness, we used a variant of the clarity measure, known as ranked list clarity \([4]\). Further details are omitted due to space constraints.

5 Results

The results of our official runs are given in Tables 1 and 2. Both the \texttt{indri05RdmT} and \texttt{indri05RdmD} runs are dependence model only runs. The \texttt{indri05RdmT} and \texttt{indri05RdmD} runs use a dependence model and mixture of relevance models with \( P(\text{bignews}) = 1, P(\text{aquaint}) = 0 \). Finally, the \texttt{indri05RdmmD} run uses the same formulation, except assumes \( P(\text{bignews}) = 0.6 \) and \( P(\text{aquaint}) = 0.4 \).

As we see, the dependence model results in a strong baseline and, when combined with mixture of relevance model expansion, produces very effective results for both title and description queries.

<table>
<thead>
<tr>
<th>Run ID</th>
<th>MAP</th>
<th>GMAP</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>indri05RdmT</td>
<td>0.2159</td>
<td>0.1354</td>
<td>1.4250</td>
</tr>
<tr>
<td>indri05RdmeT</td>
<td>0.3204</td>
<td>0.1967</td>
<td>2.3777</td>
</tr>
<tr>
<td>indri05RdmmT</td>
<td>0.3323</td>
<td>0.2061</td>
<td>2.6330</td>
</tr>
</tbody>
</table>

Table 1: Summary of Robust Track title only runs.

<table>
<thead>
<tr>
<th>Run ID</th>
<th>MAP</th>
<th>GMAP</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>indri05RdmD</td>
<td>0.1996</td>
<td>0.1015</td>
<td>0.9016</td>
</tr>
<tr>
<td>indri05RdmeD</td>
<td>0.2818</td>
<td>0.1611</td>
<td>1.9899</td>
</tr>
</tbody>
</table>

Table 2: Summary of Robust Track description only runs.

Acknowledgments

This work was supported in part by the CHR, in part by NSF grant #CNS-045018, in part by the DARPA, in part by ARDA and NSF grant #CCF-0205575, in part by NSF grant #113-0527159, and in part by SPAWAR/SYSCEN-SD grant number N66001-02-1-8903. Any opinions, findings and conclusions or recommendations expressed in this material are the authors and do not necessarily reflect those of the sponsor.

References