"A term is known by the company it keeps": On Selecting a Good Expansion Set in Pseudo-Relevance Feedback

Raghavendra Udupa[1], Abhijit Bhole[2], and Pushpak Bhattacharyya[2]

[1] Microsoft Research India
Bangalore 560080
raghavu@microsoft.com
http://research.microsoft.com/en-us/labs/india

[2] Department of Computer Science and Engineering
Indian Institute of Technology, Bombay
Mumbai 400076
{abhijit.bhole,pb}@cse.iitb.ac.in
http://www.cse.iitb.ac.in/~pb

Abstract. It is well known that pseudo-relevance feedback (PRF) improves the retrieval performance of Information Retrieval (IR) systems in general. However, a recent study by Cao et al [3] has shown that a non-negligible fraction of expansion terms used by PRF algorithms are harmful to the retrieval. In other words, a PRF algorithm would be better off if it were to use only a subset of the feedback terms. The challenge then is to find a good expansion set from the set of all candidate expansion terms. A natural approach to solve the problem is to make term independence assumption and use one or more term selection criteria or a statistical classifier to identify good expansion terms independent of each other. In this work, we challenge this approach and show empirically that a feedback term is neither good nor bad in itself in general; the behavior of a term depends very much on other expansion terms. Our finding implies that a good expansion set can not be found by making term independence assumption in general. As a principled solution to the problem, we propose spectral partitioning of expansion terms using a specific term-term interaction matrix. We demonstrate on several test collections that expansion terms can be partitioned into two sets and the best of the two sets gives substantial improvements in retrieval performance over model-based feedback.

Key words: Information Retrieval, Relevance Feedback, Pseudo-relevance Feedback, Expansion Terms, Term-Document Matrix

1 Introduction

Pseudo-relevance feedback (PRF) is a well-known method for query expansion in Information Retrieval (IR) [1]. In general, PRF uses frequent terms in the top

results of the first pass retrieval as expansion terms. The assumption underlying PRF is that the top-ranked documents contain terms related to the query terms and hence help identifying documents relevant to the query. Although conceptually simple, PRF is a very powerful technique for improving retrieval performance and is highly effective as a query expansion technique.

PRF techniques typically apply one or more criteria on the terms in the feedback documents and select terms that satisfy the criteria. Commonly employed criteria include term distributions in the feedback documents and the collection, idf, query length, and linguistic features [2]. It is common to assume that the selected expansion terms are all related to the query and use all of them in the second pass retrieval.

A recent study by Cao et al questioned the basic assumption of PRF and found that a non-negligible fraction of the expansion terms were actually harmful to the query [1] [3]. In other words, PRF would be better off if it were to use only a subset of the expansion terms. To find a good set of expansion terms, Cao et al used a SVM-based statistical classifier to select good terms. Their approach is crucially based on the assumption that *"an expansion term acts on the query independent of other terms"* and therefore, a good term can be identified independent of other expansion terms. Cao et al showed that their approach gives improvements in the retrieval performance over both model-based feedback [4] and relevance based language model [5].

In this work, we challenge the assumption that good terms can be identified independent of other expansion terms. We claim that a term is neither good nor bad in itself in general. The impact of including a term into the expansion set is a function of the term being added as well as other terms in the expansion set. The same term can behave in opposite ways depending on the company of other terms as far as retrieval performance is concerned. This is because terms interact with each other and as a consequence there is a portfolio effect. To give an analogy, if it is moderately cold, a jacket or a sweater and a shawl is more preferable than a sweater, a jacket, and a shawl together. Whereas the first two sets, i.e. {jacket} and {sweater, shawl}, are likely to keep you warm, their union will most probably make you feel uncomfortable. The effect of jacket is positive when used alone and negative when used along with sweater and shawl. In other words, the effect of jacket on your comfort depends on what other winter wear you are using along with it.

We provide solid empirical evidence for our claim: we show that in almost all topics from real test collections, a majority of the expansion terms behave inconsistently; they improve retrieval performance when paired with one set of expansion terms and degrade retrieval performance when paired with a different set.

An implication of our findings is that a good expansion set can not, in general, be discovered in a principled manner by approaches that choose terms indepen-

[1] Cao et al considered 150 topics from each of AP, WSJ, and Disk4&5 collections and 80 expansions with the largest probabilities for each topic. Approximately 50% of the terms were found to be neutral and 30% to be harmful.

dent of other terms. Such approaches make use of a flawed notion of goodness of expansion terms. **A principled solution to the problem of finding good expansion set must take into account interacting terms**. Such a solution will take a collective decision on all the expansion terms instead of independent decisions on individual terms.

We propose spectral partitioning of expansion terms as a principled solution for the problem of finding a good expansion set. Spectral partitioning takes into account interactions between terms and enables us to take a collective decision on all the expansion terms. In our partitioning experiments, we employ a weighted term-document matrix which implicitly defines a term-term interaction matrix. However, we may use any appropriately defined term-term interaction matrix in general. Given such a matrix, terms can be partitioned using standard techniques such as SVD or Graph Laplacian [6–8].

In the remainder of this paper we provide an exposition of our approach along with results of empirical investigations on multiple test collections. We start by discussing some of the important previous research work on PRF in Section 2. Next we re-examine the term independence assumption in Section 3. We describe our spectral partitioning based approach to PRF in Section 4. Next we discuss the experimental setup and results of our empirical investigations in Section 5. Finally, we discuss the results and propose some ideas for future investigation in Section 6.

2 Related work

Pseudo-Relevance Feedback has a long history in IR [1, 9]. It was first implemented in the vector space models [1] and subsequently has made its way into probabilistic models and language models [4, 5]. Since our work, like that of Cao et al [3], is in the framework of language models, we restrict our discussion to the implementations of PRF in this framework only. For an insightful and thorough discussion on feedback in language models, please see Section 5 of the recent survey on language models [10].

In the language modeling framework, documents are ranked according to the negative Kullback-Leibler (KL) divergence [11] of the query language model θ_Q with the (smoothed) document language model θ_D.

$$Score\left(Q,D\right) = -D\left(\theta_Q \| \theta_D\right) \stackrel{rank}{=} \sum_{w \in V} P\left(w\,|\theta_Q\right) \log P\left(w\,|\theta_D\right) \tag{1}$$

It is in the re-estimation of the query model θ_Q that feedback information can be leveraged. In model-based feedback [4], the original query model θ_Q is interpolated with a feedback topic model θ_F estimated from the feedback documents from the first pass retrieval:

$$P\left(w\,\middle|\theta_Q^{'}\right) = \left(1-\alpha\right) P\left(w\,|\theta_Q\right) + \alpha P\left(w\,|\theta_F\right) \tag{2}$$

where is the interpolation parameter $\alpha \in [0, 1]$ used to control the amount of feedback. There are several ways in which the topic model θ_F can be learnt

from the feedback documents in practice [4, 12]. One approach is to employ a two component mixture where one component is a fixed background model θ_C that explains the background words in the feedback documents and the other component is an unknown topic model θ_F that explains the topical words [4]. EM algorithm can be employed to estimate the topic model by maximizing the likelihood of the feedback documents. In our study, we used the feedback terms computed using this approach.

An alternative to model-based feedback is relevance-based language model, where the relevance model θ_R is estimated by assuming that the feedback documents are samples from the relevance model [5]. The original query model θ_Q is interpolated with θ_R in a manner analogous to model-based feedback. It should be noted that both model-based and relevance-based models use all the feedback terms for expansion. Whereas the topic model assigns higher probability mass to the most distinctive terms in the feedback documents, the relevance model assigns higher probability mass to the most frequent terms from the feedback documents. In contrast to both model-based feedback and relevance-based query expansion, the approach of Cao et al uses a subset of the feedback terms in expansion. They employ a statistical classifier for identifying good expansion terms [3].

3 Re-examination of the Independence Assumption

The main claim of our work is the following: **the effect of including a term into an expansion set on retrieval depends on the rest of the terms in the expansion set**. Before we go on to describe the experimental procedure for validating our claim, we discuss some motivating examples in Section 3.1.

3.1 Motivating Examples

As our first motivating example, we consider Topic 164 from TREC 3 (AP88-89): **Generic Drugs - Illegal Activities by Manufacturers**. The top 4 expansion terms of the topic model estimated from the top 10 feedback documents are **drug**, **generic**, **fda**, and **compani**. Now, consider **subcommitte** and **result**, two candidate expansion terms for this topic. According to term goodness criterion of Cao et al (see Section 3 of [3]), which makes independence assumption, **subcommitte** is a bad term whereas **result** is a good term. However, they behave very differently with different expansion sets. For instance, **subcommitte** acts as a good term when used with the set {**drug**, **generic**} and when used with {**drug**, **generic**, **fda**}, it acts as a bad term. Similarly, **result** acts as a good term with {**drug**, **generic**, **fda**} and becomes a bad term when used with {**drug**, **generic**, **fda**, **compani**}.

As our second example, we consider the title of Topic 308 from TREC 6 (Disk4&5): **Implant Dentistry**. The top 4 expansion terms of the feedback topic model are **devic**, **implant**, **chiropract**, and **fda**. Consider **prosthesi** and **requir**, two candidate expansion terms for this topic. According to term

goodness criterion of Cao et al **prosthesi** is a good term whereas **requir** is a bad term. However, **prosthesi** acts as a good term when used with the set {**devic**, **implant**} and when used with {**devic**, **implant**, **chiropract**}, it acts as a bad term. On the other hand, **requir** acts as a bad term when used with the set {**devic**, **implant**, **chiropract**} and as a good term when used with {**devic**, **implant**, **chiropract**, **fda**}.

Continuing the investigation of the above topics, we did the following experiment: For each feedback term t, we checked the effect of adding t to each of the sets $T_1, \ldots, T_{25}$, where T_k is the set of top k terms of the feedback topic model [2]. Our aim was to find out how many feedback terms behave consistently with respect to the sets $T_1, \ldots, T_{25}$. If the term independence assumption were indeed valid then most terms must behave consistently. However, our investigation revealed that 56% of the feedback terms of the query **Generic Drugs - Illegal Activities by Manufacturers** (Topic 164, TREC 3) are inconsistent. In the case of the query **Implant Dentistry** (Topic 308, TREC 6), the percentage of inconsistent feedback terms was even higher, 92%.

Figure 3.1 shows the behavior of the term **prosthesi** (Topic 308, TREC 6) when it is used with $T_1, \ldots, T_{25}$. As can be seen from the figure, the behavior of the term is highly inconsistent across the expansion sets. It acts as good, bad, or neutral depending on the expansion set with which it is used. These examples

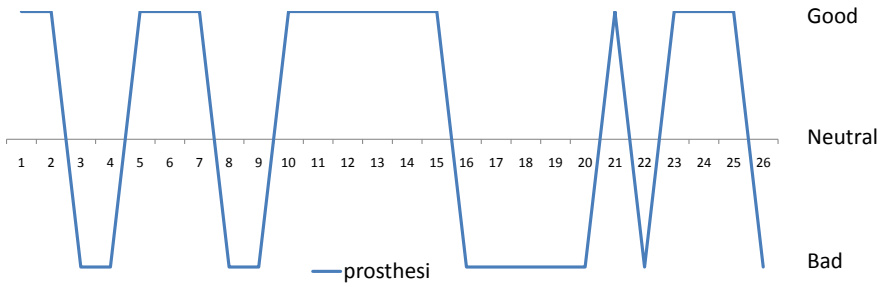


Fig. 1. The inconsistent behavior of **prosthesi** when used with different expansion sets

not only show that terms are neither good nor bad in isolation but also suggest that term selection strategies that make independence assumption must not be trusted in general.

[2] We ranked the feedback terms according to $P(w|\theta_F)$.

3.2 Empirical Validation of the Term Dependence Claim

In this section, we describe the experiments we did on topics from several collections to determine the set of inconsistent terms (i.e. terms which behave differently with different expansion sets) from each topic.

Model Let Q be a topic and θ_F be the feedback topic model estimated using the two component mixture model approach described in Section 2. Let S_F be the set of top 100 terms in θ_F and $S \subseteq S_F$ be an expansion set. We formed a new feedback topic model θ_F^S as follows:

$$P\left(w\left|\theta_F^S\right.\right) = \begin{cases} \frac{P\left(w\left|\theta_F\right.\right)}{\sum_{w' \in S} P\left(w'\left|\theta_F\right.\right)} \text{ if } w \in S \\ 0 \text{ otherwise} \end{cases} \tag{3}$$

The re-estimated query model is then $P\left(w\left|\theta_Q^{'}\right.\right) = (1-\alpha)\, P\left(w\left|\theta_Q\right.\right) + \alpha P\left(w\left|\theta_F^S\right.\right)$. To integrate a new term t to θ_F^S, we used the following scoring function:

$$Score\left(Q, D\right) = \sum_{w \in V} P\left(w\left|\theta_Q^{'}\right.\right) \log P\left(w\left|\theta_D\right.\right) + \eta \log P\left(t\left|\theta_D\right.\right) \tag{4}$$

where $\eta > 0$ is the weight for the term t. In our experiments, α was set to 0.2 and η was set to 0.05 which is in the same range as the weights of the top 10 terms of the feedback model θ_F.

Let $MAP\left(Q \cup S\right)$ be the retrieval performance when S is the expansion set and $MAP\left(Q \cup S \cup t\right)$ be the retrieval performance after adding t to S with weight η. We can now define the relative marginal gain in retrieval performance as follows:

$$RMG\left(S, t\right) = \frac{MAP\left(Q \cup S \cup t\right) - MAP\left(Q \cup S\right)}{MAP\left(Q \cup S\right)} \tag{5}$$

Using relative marginal gain, we labeled t as follows: good if $RMG\left(S, t\right) \geq \delta$, bad if $RMG\left(S, t\right) \leq -\delta$ and neutral otherwise. Here $\delta > 0$ is a cutoff which in our experiments was set to 0.005.

Testing for Inconsistency Let T_k denote the set of top k terms of the feedback model θ_F and t be a term [3]. For each of the expansion sets $T_1, \ldots, T_{25}$, we assigned a label $l_k \in \{\text{good, bad neutral}\}$ to t depending on the effect of adding t to T_k. We call t consistent if all the labels l_k, $k = 1, \ldots, 25$ are identical and inconsistent otherwise. A consistent term is one which behaves the same way with each of the expansion sets $T_1, \ldots, T_{25}$. Using this procedure we estimated N_Q, the number of inconsistent terms in each topic.

[3] Model-based feedback produces a large number of feedback terms. We used only the top 100 terms.

Test Results We used several collections in our study: CLEF 2000-02, CLEF 2003,05,06 , AP (Associated Press 88-89, TREC Disks 1 and 2), WSJ (Wall Street Journal , TREC Disks 1&2), SJM (San Jose Mercury, TREC Disk 3) and TREC Disks 4&5 (minus the Congressional Record). Table 1 shows the mean and standard deviation of the number of inconsistent terms in the topics for several test collections and Figure 2 shows the histogram of inconsistent terms for the AP collection. We observed that very few topics have a small number of inconsistent terms. For most topics in the collections, 30-60 terms out of the top 100 are inconsistent and about 50% of the terms are inconsistent on an average. These statistics unequivocally tell that inconsistency is a major problem and can not be ignored while selecting terms.

In a different experiment, we used the relevant documents of a topic to estimate its topic model. This was to verify whether the source of the inconsistent terms in PRF was non-relevant documents in the feedback. We repeated the inconsistency test with the top 100 terms of the topic model estimated directly from the relevant documents. Table 2 shows the mean and standard deviation of the number of inconsistent terms in the topics for several test collections. We observed that the mean and standard deviation were similar to those in the previous experiment. This means that even if relevant documents are provided as feedback, inconsistency remains an important issue.

Table 1. Inconsistent Terms (with top 10 documents as feedback)

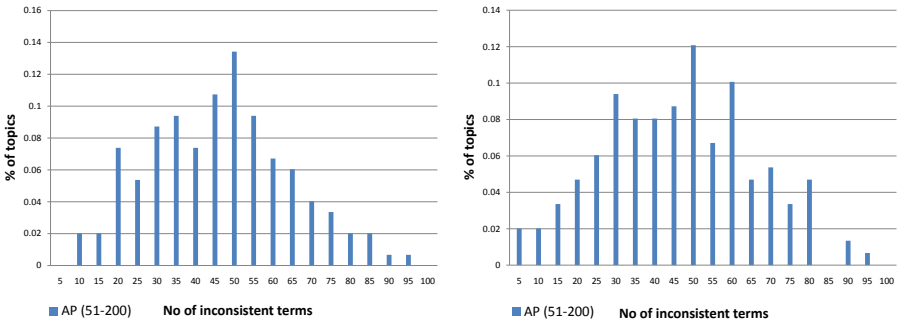
Collection	Mean	Std.Dev.
CLEF(1-140)	58.13	21.24
CLEF(141-200,251-300)	49.61	21.06
AP(51-200)	44.37	18.13
WSJ(51-200)	50.69	18.46
SJM(51-150)	51.46	20.43
Disk 4&5(301-450)	48.62	18.41

Table 2. Inconsistent Terms (with relevant documents as feedback)

Collection	Mean	Std.Dev.
CLEF(1-140)	52.42	25.22
CLEF(141-200,251-300)	49.35	23.44
AP(51-200)	44.21	19.66
WSJ(51-200)	50.62	18.90
SJM(51-150)	51.28	21.83
Disk 4&5(301-450)	47.06	18.83

4 Finding Good Expansion Set by Spectral Partitioning

We now describe a principled approach for finding good expansion sets that takes into account term interactions. The key idea here is to form a weighted term-document matrix and partition it into two sets using Singular-Value Decomposition (SVD) [7]. Geometrically, the principal singular vector of the term-document matrix gives the direction that captures most of the spread (or variance) in the data. In pattern recognition and machine learning literature, this is also known as principal components analysis and is known to provide a very good low-dimensional representation of the data. Spectral partitioning techniques are very effective in practice and have been used successfully in many applications



(a) Using top 10 feedback documents (b) Using relevant documents

Fig. 2. Distribution of N_Q in the AP collection

[6, 13]. Further, they are highly useful in understanding global properties of a phenomenon using local interactions. Note that the covariance matrix obtained by multiplying the (centered) term-document matrix with its transpose captures pair-wise interactions. Higher powers of the covariance matrix in turn capture higher order interactions.

4.1 Partitioning Algorithm

Let A be a matrix whose rows represent the candidate feedback terms [4] $\{t_i\}_{i=1}^{m}$ and the columns represent the feedback documents $\{D_j\}_{i=1}^{n}$. Let $[A]_{ij} = a_{ij}$ be a measure of the interaction between term t_i and document D_j. We express a_{ij} as $a_{ij} = \text{global}\,(t_i) * \text{local}\,(t_i, D_j)$ where $\text{global}\,(t_i)$ is a global weighting function and $\text{local}\,(t_i, D_j)$ is a local weighting function.

The global weighting function measures the informativeness of terms with respect to the feedback documents. Our choice for this function is the following:

$$\text{global}\,(t) = \ln\left(\frac{n_t}{n} \Big/ \frac{df_t}{N}\right) \tag{6}$$

where n (and resp. N) is the number of feedback documents (and resp. number of documents in the collection) and n_t (and resp. df_t) is the document frequency of t in the feedback corpus (and resp. document frequency of t in the collection). We can write

$$\text{global}\,(t) = \ln\frac{N}{df_t} - \ln\frac{n}{n_t} = idf_t - idf'_t \tag{7}$$

where $idf'_t = \ln\frac{n}{n_t}$ is the idf of t in the feedback corpus. It can be easily shown that $\text{global}\,(t) \leq \ln\frac{N}{n}$ with equality holding only when $n_t = df_t$. Thus, according

[4] Candidate expansion terms are those terms from the feedback documents whose $idf > \ln 10$ and collection frequency ≥ 5. When there are more than 100 such terms we take the top 100 according to their frequency in the feedback documents.

to the global weighting function, a term t is more informative than another term s if $idf_t - idf'_t > idf_s - idf'_s$ or equivalently if $idf_t - idf_s > idf'_t - idf'_s$. Therefore, from the point of view of PRF, t can be more informative than s even when $idf_t < idf_s$. Finally, our choice for local (t_i, D_j) is $P(t|D)$, the smoothed unigram probability of the term t in document D [14].

We center the matrix A such that the mean of the row vectors is the $\vec{0}$ vector. The Singular Value Decomposition (SVD) of the term-document matrix A is then the following:

$$A = U\Sigma V^T \tag{8}$$

where U and V are orthogonal matrices and Σ is a diagonal matrix.

The sign of the terms in the principal left singular vector $\vec{u_1}$ suggests a principled way to partition the terms into two sets [7]. We form the first set, S^+ by taking those terms whose sign is positive in the principal singular vector. Similarly, we form the second set, S^- by taking those terms whose sign is negative. We remove terms from S^+ and S^- that have an absolute weight below a threshold.

4.2 From partitions to feedback model

There are several ways in which we can form a feedback model using S^+ and S^-. In our experiments we used two very simple methods as our goal was to mainly validate the goodness of the expansion sets produced by spectral partitioning. In the first method (SU), we assigned uniform probability to all the terms in the set and in the second (SF), we assigned a probability proportional to the frequency of the term in the feedback documents. In both methods, we formed feedback models θ_F^+ (which allocates non-zero probability to only terms of S^+) and θ_F^- (which allocates non-zero probability to only terms of S^-). The two models θ_F^+ and θ_F^- represent two different choices for expansion. As our goal in this study was to demonstrate that spectral partitioning separates the good set of terms from the bad, we used the one that gave the best results for the topic.

5 Experimental Results

We tested our spectral partitioning idea on the following test collections: CLEF 2000-02, CLEF 2003,05,06 , AP (Associated Press 88-89, TREC Disks 1 and 2), WSJ (Wall Street Journal , TREC Disks 1&2), SJM (San Jose Mercury, TREC Disk 3) and TREC Disks 4&5 (minus the Congressional Record). We used two baselines, language model (LM) with two stage Dirichlet smoothing and the mixture feedback model (MF) [4]. We interpolated the feedback model with a weight of 0.5. We stemmed the words using the well-known Porter stemmer and removed stop-words from topics and documents.

Table 3 shows the retrieval results for various models. We did t-test to determine the significance of the results. Both SF and SU gave substantially better results than the LM and MF baselines on all test collections. We observed $> 15\%$

improvement in MAP over LM for all the test collections. The spectral methods had substantially better P@10 compared to both LM and MF. This means that spectral expansion was able to retrieve relatively larger number of relevant results in the top 10. Further, we observed improvement in all performance metrics. Finally, the performance of SU was comparable with that of SF which means that the expansion sets are robust to perturbations in weights.

Table 3: Retrieval results (** = significant at $p < 0.01$ over LM) (ˆˆ = significant at $p < 0.01$ over MF)

Collection	Model	P@10	P@100	MAP	% Improvement	Recall
CLEF (1 - 140)	LM	0.3828	0.1181	0.4338	-	0.8936
	MF	0.4148	0.1310	0.4417	1.82%	0.9348
	SU	0.4459	0.1383	0.5037	16.11% **ˆˆ	0.9578
	SF	0.4484	0.1398	0.503	15.95% **ˆˆ	0.9632
CLEF (141 - 200, 251 - 300)	LM	0.3684	0.1573	0.3808	-	0.8172
	MF	0.3836	0.1694	0.4053	6.43% **	0.9239
	SU	0.4296	0.1737	0.4516	18.59% **ˆˆ	0.9357
	SF	0.4362	0.1739	0.4474	17.49% **ˆˆ	0.9345
AP (51 - 200)	LM	0.4450	0.2655	0.2772	-	0.6504
	MF	0.4732	0.3026	0.327	17.97% **	0.7217
	SU	0.5201	0.3303	0.3641	31.35% **ˆˆ	0.7565
	SF	0.5315	0.3312	0.3648	31.60% **ˆˆ	0.7564
WSJ (51 - 200)	LM	0.4573	0.2611	0.2660	-	0.6438
	MF	0.4773	0.2884	0.3027	13.79%**	0.6971
	SU	0.5193	0.2975	0.3238	21.73% **ˆˆ	0.7106
	SF	0.5113	0.2969	0.3213	20.79% **ˆˆ	0.7173
SJM (51 - 150)	LM	0.3043	0.1572	0.2074	-	0.6173
	MF	0.3234	0.1736	0.2350	13.31%**	0.6773
	SU	0.3649	0.1832	0.2601	25.41% **ˆˆ	0.6916
	SF	0.3691	0.1816	0.263	26.81% **ˆˆ	0.6992
Disks 4& 5 (301 - 450)	LM	0.4247	0.1987	0.2275	-	0.5359
	MF	0.4360	0.2152	0.2505	10.11% **	0.5746
	SU	0.4693	0.2385	0.2848	25.19% **ˆˆ	0.6153
	SF	0.4707	0.2413	0.2876	26.42% **ˆˆ	0.6205

Table 4 shows the expansion sets for three topics. Firstly, we observe that the terms in each set are topically coherent. Consider Topic 311 of TREC for instance. All the expansion terms found by our Spectral Partitioning algorithm are topically related to the query **Industrial Espionage**. In Topic 112 of CLEF, we observe that the expansion terms found by our Spectral Partitioning algorithm include the names of the director and lead actor of the movie **Pulp Fiction**. Finally, in Topic 63 of AP too, the selected expansion terms are topically related to **Machine Translation**.

Table 4. Expansion set using Spectral Partitioning

Pulp fiction (Topic 112, CLEF)		**Machine Translation** (Topic 63, TREC 3)		**Industrial Espionage** (Topic 311, TREC 6)	
Term	$P(t\,\vert\theta_F)$	Term	$P(t\,\vert\theta_F)$	Term	$P(t\,\vert\theta_F)$
movi	0.25	comput	0.28	vw	0.23
film	0.20	english	0.21	gm	0.17
travolta	0.20	word	0.14	german	0.14
tarantino	0.18	languag	0.09	lopez	0.13
actor	0.09	human	0.07	investig	0.13
quentin	0.06	recogn	0.06	opel	0.07
cann	0.02	voic	0.05	motor	0.07
		pen	0.05	volkswagen	0.05
		dictionari	0.04	wolfsburg	0.03

6 Conclusion

We showed that the term independence assumption for the selection of good expansion terms does not hold in practice through a thorough study of the effect of an expansion term in the presence of other terms. In practice, about 50% of the expansion terms are inconsistent, i.e. they behave differently with different expansion sets. Our empirical finding implies that good expansion sets can not be discovered by methods that make independence assumption in general. As a principled method of discovering good expansion sets, we proposed spectral partitioning of term-term interaction matrix which takes into account term interactions of all orders. We demonstrated that the expansion sets produced by spectral partitioning give substantially better retrieval results than both language model and model-based feedback on topics from several test collections.

In a future study, we will explore more sophisticated methods for forming the feedback model from the partitions produced by our method. For instance, we can leverage weights of terms in the principal left singular vector while forming the feedback model. Another direction of research is how to choose between θ_F^+ and θ_F^- for a given topic.

References

1. Buckley, C., Salton, G., Allan, J.: Automatic retrieval with locality information using smart. In: TREC. (1992) 59–72
2. Carpineto, C., Romano, G.: Towards more effective techniques for automatic query expansion. In: ECDL '99: Proceedings of the Third European Conference on Research and Advanced Technology for Digital Libraries, London, UK, Springer-Verlag (1999) 126–141
3. Cao, G., Nie, J.Y., Gao, J., Robertson, S.: Selecting good expansion terms for pseudo-relevance feedback. In: SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, ACM (2008) 243–250

4. Zhai, C., Lafferty, J.: Model-based feedback in the language modeling approach to information retrieval. In: In Proceedings of Tenth International Conference on Information and Knowledge Management. (2001) 403–410
5. Lavrenko, V., Croft, B.W.: Relevance based language models. In: SIGIR '01: Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, ACM Press (2001) 120–127
6. Chung, F.R.K.: Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92) (Cbms Regional Conference Series in Mathematics). American Mathematical Society (February 1997)
7. Meyer, C., Basabe, I., Langville, A.: Clustering with the svd. In: Workshop on Numerical Linear Algebra, the Internet and its Applications. (Monopoli, 2007)
8. von Luxburg, U.: A tutorial on spectral clustering. Technical Report 149, Max Planck Institute for Biological Cybernetics (August 2006)
9. Efthimiadis, E.N.: Query expansion. Annual Review of Information Systems and Technology **31** (1996) 121–187
10. Zhai, C.: Statistical language models for information retrieval a critical review. Found. Trends Inf. Retr. **2**(3) (2008) 137–213
11. Kullback, S., Leibler, R.A.: On information and sufficiency. The Annals of Mathematical Statistics **22**(1) (1951) 79–86
12. Tao, T., Zhai, C.: Regularized estimation of mixture models for robust pseudo-relevance feedback. In: SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, ACM (2006) 162–169
13. Spielman, D.A., Teng, S.: Spectral partitioning works: Planar graphs and finite element meshes. Technical report, Berkeley, CA, USA (1996)
14. Zhai, C., Lafferty, J.: A study of smoothing methods for language models applied to information retrieval. ACM Trans. Inf. Syst. **22**(2) (2004) 179–214
15. Lafferty, J., Zhai, C.: Document language models, query models, and risk minimization for information retrieval. In: SIGIR '01: Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, New York, NY, USA, ACM (2001) 111–119
16. Robertson, S.E.: On term selection for query expansion. J. Doc. **46**(4) (1990) 359–364
17. Smeaton, A.F., van Rijsbergen, C.J.: The retrieval effects of query expansion on a feedback document retrieval system. Comput. J. **26**(3) (1983) 239–246
18. Robertson, S.E., Jones, S.K.: Relevance weighting of search terms. Journal of the American Society for Information Science **27**(3) (1976) 129–146