

Compact Aspect Embedding for Diversified Query Expansions

Xiaohua Liu, Arbi Bouchoucha, Alessandro Sordoni and Jian-Yun Nie

Dept. of Computer Science and Operations Research
University of Montreal
Montreal (Quebec), Canada
{liuxiao, bouchoar, sordonia and nie}@iro.umontreal.ca

Abstract

Diversified query expansion (DQE) based approaches aim to select a set of expansion terms with less redundancy among them while covering as many query aspects as possible. Recently they have experimentally demonstrate their effectiveness for the task of search result diversification. One challenge faced by existing DQE approaches is to ensure the aspect coverage. In this paper, we propose a novel method for DQE, called compact aspect embedding, which exploits trace norm regularization to learn a low rank vector space for the query, with each eigenvector of the learnt vector space representing an aspect, and the absolute value of its corresponding eigenvalue representing the association strength of that aspect to the query. Meanwhile, each expansion term is mapped into the vector space as well. Based on this novel representation of the query aspects and expansion terms, we design a greedy selection strategy to choose a set of expansion terms to explicitly cover all possible aspects of the query. We test our method on several TREC diversification data sets, and show that our method significantly outperforms the state-of-the-art search result diversification approaches.

Introduction

The short and ambiguous nature of the Web queries (Agrawal et al. 2009) represents a main challenge for a search engine to understand the user intent. Search result diversification (SRD) has been proposed to tackle this challenge by returning a list of relevant results covering as many aspects of the query as possible. SRD can be implemented by selecting or re-ranking the initial search results, taking into account the diversity among them (Carbonell and Goldstein 1998; Zhai, Cohen, and Lafferty 2003; Zhu et al. 2007). Some recent studies (Bouchoucha, He, and Nie 2013) find that the initial search results may not be diverse enough to represent all the aspects of the query. Therefore, a pre-retrieval query expansion can be conducted to expand the query in different directions, which motivates recent studies on diversified query expansion (DQE) for SRD (Bouchoucha, He, and Nie 2013; Bouchoucha, Liu, and Nie 2014). Experimentally, DQE based approaches, e.g.,

the one presented by Bouchoucha et al. (2013), outperform the state-of-the-art existing SRD methods on TREC data, including those based on pseudo-relevance feedback (PRF)(Agrawal et al. 2009; Santos, Macdonald, and Ounis 2010).

A typical DQE approach consists of three steps. It first generates a set of expansion term candidates using one or several external resources, e.g., ConceptNet (Speer and Havasi 2012), Wikipedia, query logs, or initial feedback documents. Then it selects a set of diverse expansion terms from the candidate expansion terms, following some principled method. Finally, it combines expansion terms and the original query into one extended query (in which each term has a weight) and uses that query to obtain a set of diversified search results. As aspects of a query are not explicitly specified in a realistic situation, existing DQE approaches assume an aspect of the query can be represented by one or several expansion terms. Their focus is then on removing the redundancy among expansion terms, while no attention has been paid on how well the selected terms can cover all the query aspects. For example, in the work of Bouchoucha et al. (2013), they define two functions to measure the similarities between two expansion terms and between an expansion term and the query, respectively, and based on the two similarity functions they obtain diversified expansion terms by following the Maximal Marginal Relevance (MMR) (Carbonell and Goldstein 1998) principle. Notice that their approach cannot guarantee that the expansion terms selected cover all aspects of the query. Indeed, we have observed that some expansion terms are ignored, although they cover some new aspects of the query. This is because they cover some aspects already covered by the expansion terms already selected.

We propose a novel method called compact aspect embedding that explicitly extracts query aspects. In particular, we represent each expansion term with a vector, and use the span of the expansion term vectors to represent the aspect space of the query. Each vector of a basis of the aspect space could be considered as an aspect. Since a query often has a limited number of aspects, we are motivated to learn a compact vector space - a vector space of low rank, which is achieved by using trace norm regularization

(Ji and Ye 2009). Observing that the expansion terms of a query may be organized into several independent groups¹, and each group may have a different degree of relevance to the original query, we choose one special basis consisting of the eigenvectors²: each eigenvector denotes an aspect, called aspect vector hereafter, and the absolute value of the corresponding eigenvalue is considered as the association strength of the aspect with the query, called aspect weight hereafter.

Based on this vector representation for an expansion term and an aspect, we design the following greedy selection strategy to obtain a set of expansion terms that tries to explicitly cover all aspects: we process the aspect vectors by their weights in descending order by selecting several new expansion terms (not selected before) that are not similar to the expansion terms already selected and are also close to the aspect vector under consideration in the aspect space. Note that, compared with latent Dirichlet allocation (LDA) (Blei, Ng, and Jordan 2003), which also does not need to know the explicit sub topics of the given query, our method exploits supervised learning to learn the aspect vectors and expansion term vectors, and does not enforce a probabilistic interpretation of the learnt vectors (which will be explained in Section 3).

We evaluated our approach using TREC 2009, 2010 and 2011 Web tracks. The experimental results confirm that our method yields better results than the previous DQE approaches and other typical SRD approaches.

We make several contributions in this paper. 1) we propose to use a low rank vector space to represent the potential intents of a query; 2) we present a greedy expansion term selection strategy based on the learnt aspect vector space aiming to cover all aspects; and 3) we experimentally show the effectiveness of our method on TREC 2009, 2010 and 2011 Web tracks.

The remainder of this paper is organized as follows. Section 2 overviews related work. Section 3 presents details of our approach. Section 4 describes our experiments. Finally in Section 5, we conclude with future work.

Related Work

Implicit SRD. Existing SRD methods can be roughly classified into two categories: implicit SRD and explicit SRD. An implicit SRD approach considers the relations among documents to generate ranked lists that convey both relevant and diverse information about a query. Maximal Marginal Relevance (MMR) (Carbonell and Goldstein 1998) is an early example of implicit SRD, which selects documents that maximize relevance and reduce redundancy to higher ranked documents. Zhai et al. propose a probabilistic version of MMR (2003). Zhu et al. approached the diversi-

fication problem as an absorbing Markov chain using random walks to prevent redundant items from receiving a high rank by turning ranked items into absorbing states (2007), which will decrease the importance of items that are similar to them, thereby promoting items that are dissimilar to them.

Explicit SRD. An explicit SRD approach uses a set of terms, manually defined or automatically extracted, to explicitly represent the aspects of the query, and tries to cover as much as possible the different aspects of the query. Several resources have been exploited to define or extract query aspects. For example, the explicit Query Aspect Diversification framework (xQuAD) (Santos, Macdonald, and Ounis 2010) extracts query reformulations from three major Web search engines to represent query aspects. A document is re-ranked according to how it can cover the uncovered aspects. Dang and Croft (2012) use the official sub-topics manually identified by TREC assessors and query suggestions provided by a commercial search engine as aspects, and introduce PM-2, a two stage framework which first determines the aspect that best maintains the overall proportionality for each position in the result ranking list, and then selects the best document on that aspect. In their recent study (2013), Dang and Croft extend PM-2 by using the topic terms extracted with an algorithm for document summarization from feedback documents as query aspects, with the hope that the expanded query can cover more query aspects. Dou et al. (2011) propose a multi-dimensional SRD framework that exploits four data sources, including anchor texts, query logs, search result clusters and web sites to mine query subtopics on multiple dimensions. Such subtopics are used to diversify documents by considering both their relevance and their novelty, following the MMR principle (Carbonell and Goldstein 1998). He et al. (2012) propose the combination of click logs, anchor text and web n-grams to generate related terms for query expansion. They then use a graph to organize these terms, and compute the similarities between two suggested terms by using random walk. They experimentally show that with suggested queries, their approach improves diversification performance. Note that, their approach selects expansion terms according to their similarity to the query terms, and does not consider the possible redundancy among expansion terms.

DQE. Diversified query expansion (DQE) represents recent efforts in explicit SRD, with the direct goal of obtaining a set of diversified expansion terms. Note that, different from general query expansion, DQE aims at improving search result diversification: not only to cover more relevant documents, but also to cover more diversified documents. Therefore, the diversity of the expansion terms should be explicitly considered. Several approaches have been proposed for DQE. Vargas et al. (2013) propose the first DQE approach, which employs xQuAD to select diverse expansion terms from feedback documents. Bouchoucha et al. (2013) use an external resource - ConceptNet to get a set of diverse expansion terms. In their recent work (2014), they present a general framework that integrates multiple resources to generate diverse expansion terms according to a principle similar to MMR (Carbonell and Goldstein 1998). Although term dissimilarity is a criterion used in the selection, there is no

¹This is verified by a lot of clustering based approaches to SRD (He, Meij, and de Rijke 2011; Nguyen and Hiemstra 2012).

²The rank of a matrix equals to the number of non-zero eigenvalue of that matrix. For a real matrix, which is the case in our study, eigenvectors corresponding to different non-zero eigenvalues are linearly independent to each other. For each non-zero eigenvalue, we choose one eigenvector.

explicit representation of query aspects. We argue that an explicit representation of query aspects may be important for DQE: one would be able to estimate not only if two expansion terms are dissimilar, but also if they are related to different aspects. This latter factor is important for the generation of diversified queries covering all the query aspects. To the best of our knowledge, this work is the first one to explicitly represent the aspects of the query using a low rank vector space. Based on this representation we further present a greedy selection strategy to choose a set of diverse terms that cover all the aspects of the query. This is in sharp contrast with the existing DQE approaches, which employ terms or topical models to represent the aspects of the query.

Embedding. Embedding is an abstract representation created to represent latent semantics. With embedding, any object (e.g. a term, an aspect, etc.) can be mapped to a vector in the embedding space, thus has a latent semantic representation. Recently, the idea of embedding has been successfully exploited for a wide range of tasks. For example, Koren et al. exploit matrix factorization technologies to map users and movies to the same vector space, and win Netflix Prize competition (2009). Huang et al. use a multi-layer neural network to learn vector representations for the document using click-through data, which outperforms other latent semantic models for the Web document ranking task (2013). He et al. use LDA to obtain a vector for each document, which models the distribution of latent topics of that document (2012). In this work, we use trace norm regularization (Ji and Ye 2009; Cai, Candès, and Shen 2010) to learn a low rank vector space to represent the aspects of the query, which encodes our prior knowledge that a query often has a limited number of aspects.

Compact Aspect Embedding

Overview. Our method is an instance of DQE, and consists of three steps. Given a query, we first generate expansion terms using some external resource, and define a function to measure the similarity between two expansion terms, following the studies of Bouchoucha et al. (2013; 2014). In our work, we use search logs, which has been widely adopted by recent SRD related studies. Then we map expansion terms into a low-rank vector space. With the learnt vector space, we select an eigenvector (aspect vector) for each non-zero eigenvalue to represent an aspect of the query in the vector space. Accordingly, we use the absolute value of the eigenvalue (aspect weight) to model the associations strength of the corresponding aspect with the query. To ensure that the expansion terms selected are relevant to the query and cover all the aspects of the query, we design the following greedy selection strategy: we first order the aspect vectors in descent order by their weights; and then for each aspect vector we select several expansion terms that cover this aspect while not being redundant with already selected expansion terms. Following (Dang and Croft 2012), we consider the proportionality of an aspect, and make the number of selected expansion terms for an aspect proportional to the weight of the aspect. Finally, the selected expansion terms

and the original query terms are combined to formulate a new query, which is submitted to a retrieval system, e.g., Indri³. We take the returned search results as the diversified search results, following Bouchoucha et al. (2013). It is the second step that makes our method different from other related approaches, which we now discuss in detail.

Embedding Framework. Similar to latent Dirichlet allocation (LDA) (Blei, Ng, and Jordan 2003), our embedding framework does not need to know the explicit sub topics of the given query, and attempts to learn a vector for each expansion term. Furthermore, it uses a vector to represent an aspect of the query as well, and all the aspect vectors span a low-rank vector space, in which we can consistently measure the distance between two expansion terms and the distance between an expansion term and an aspect. This is different from LDA, in which an aspect is modelled as a topic, i.e., one dimension of a topical vector. Another noticeable difference is that our embedding framework uses supervised learning to learn the low-rank aspect space, by enforcing that similar expansion terms should be close in the aspect space.

We introduce the following notations before we formally define our embedding framework: q , the given query; E , the expansion terms related to q ; $e \in E$, an expansion term; \vec{e} , the column vector corresponding to expansion term e ; \mathbf{E} , the matrix with each column representing an expansion term vector, which also denotes the vector space we want to learn; $\|\cdot\|_F$, the Frobenius-norm of a matrix (resp. a vector), defined as the sum of the absolute squares of all elements of the matrix (resp. a vector); $\|\cdot\|_*$, the trace norm of a matrix, defined as the sum of the singular values of the matrix; \mathbf{E}^T , the transpose of matrix \mathbf{E} ; $\mathbf{S} = (s_{ij})$, the similarity matrix, where s_{ij} denotes the similarity between two expansion terms e_i and e_j (which will be explained soon).

With the above notations, we define the following optimization problem:

$$\begin{aligned} \min & \frac{1}{2} \|\mathbf{E}^T \mathbf{E} - \mathbf{S}\|_F^2 + \eta \|\mathbf{E}\|_* \\ \text{subject to: } & \|\vec{e}\|_F^2 = 1, \forall e \in E. \end{aligned} \quad (1)$$

Formula 1 clearly shows that we are seeking for an aspect vector space \mathbf{E} with two desirable characteristics: 1) similar expansion terms should be close in the aspect vector space; and 2) the vector space should have a low rank, which is implemented by enforcing the trace norm regularization. η controls the trade-off. Note that directly obtaining a low rank matrix in general is NP-hard due to the combinatorial nature of the rank function (Ji and Ye 2009). The trace norm is a commonly-used convex relaxation of the rank function, since it is the convex envelope of the rank function over the unit ball of spectral norm. Indeed, under certain conditions, the low rank solution can be recovered exactly via minimizing the trace norm (Recht, Xu, and Hassibi 2008). In practice, the trace norm relaxation has been shown to yield low-rank solutions and it has been used widely in many scenarios (Tomioka and Aihara 2007; Ma, Goldfarb, and Chen 2011).

³<http://www.lemurproject.org/indri.php>

We use the singular value thresholding algorithm (Cai, Candès, and Shen 2010) to solve this problem. This algorithm is easy to implement, and is effective both in terms of computational cost and storage requirement when the minimum nuclear-norm solution is also the lowest-rank solution.

Once we obtain \mathbf{E} , we get a vector for each expansion term $e \in E$. We further compute the eigenvalues $\sigma_i, i = 1, 2, \dots, N$, where N is the rank of \mathbf{E} . For each eigenvalue σ_i , we choose one normalized eigenvector, denoted by $\vec{\alpha}_i$, where $\|\vec{\alpha}_i\|_F^2 = 1$. Note that $\{\vec{\alpha}_i, i = 1, 2, \dots, N\}$ form a basis of \mathbf{E} . Therefore, we use $\vec{\alpha}_i$ to represent the i^{th} aspect, and use $|\sigma_i|$ to represent its weight, which indicates to which degree the aspect is related to the query q . Hereafter, we assume $\vec{\alpha}_i, i = 1, 2, \dots, N$ are in descending order by their weights.

Expansion Terms Generation. We use query logs to extract expansion terms. We consider two kinds of queries as expansion terms: 1) the queries that share the same click-through data with q , and 2) the reformulated queries of q that appear in a user session within a time window of 30 minutes. For a query, we select at most K (experimentally set to 100) expansion terms.

For each pair of expansion terms e_i and e_j , we compute their similarity, i.e., s_{ij} , using the Jaccard distance by considering the number of the expansion terms that include both e_i and e_j :

$$s_{ij} = \frac{|Q_i \cap Q_j|}{|Q_i \cup Q_j|} \quad (2)$$

where $Q_i = \{e | e \in E, e_i \in q\}$ (resp. $Q_j = \{e | e \in E, e_j \in q\}$) is the subset of E that includes e_i (resp. e_j).

Expansion Term Diversification. Expansion terms generated from query logs often include terms that are similar to each other. Therefore, we design Algorithm 1, a greedy selection strategy, to obtain a global list of diversified expansion terms, with the goal of removing redundancy among expansion terms and covering as many aspects of the original query as possible. For each aspect vector, we decide the number of expansion terms to be selected for the aspect, following Formula 3:

$$K_i = \lceil 2 \cdot N \cdot \frac{|\sigma_i|}{\sum_{j=1, \dots, N} |\sigma_j|} \rceil \quad (3)$$

where K_i denotes the number of expansion terms we will choose for the i^{th} aspect. This formula ensures two things: 1) an aspect with higher weights may have more expansion terms; and 2) each aspect will have at least one expansion term. For each aspect under consideration, a Maximal Marginal Relevance-based Expansion (MMRE) (Bouchoucha, Liu, and Nie 2014) is performed to choose an expansion term e^* that covers that aspect and is not similar to the expansion terms already selected, as defined in Formula 4. The selection process stops if no expansion term candidate is available or for every aspect the required number of expansion terms have been selected.

$$e^* = \operatorname{argmax}_{e \in E - ES} \{\beta \cdot \operatorname{sim}(\vec{e}, \vec{\alpha}_i) - (1 - \beta) \cdot \max_{\vec{e}' \in ES} \operatorname{sim}(\vec{e}, \vec{e}')\} \quad (4)$$

Algorithm 1 Query Term Diversification.

Require: Expansion terms E ; aspect vector space \mathbf{E} ; aspect vectors $\vec{\alpha}_i, i = 1, 2, \dots, N$.

- 1: Initialize ES , the set of selected expansion terms, to empty: $ES = \emptyset$.
- 2: Initialize k_i , the expansion terms already selected for the i^{th} aspect, to 0: $k_i = 0, i = 1, \dots, N$.
- 3: **while** $E \neq \emptyset$ and $k_j < K_j, j = 1, \dots, N$ **do**
- 4: **for** $i = 1, 2, \dots, L$ **do**
- 5: **if** $E == \emptyset$ **then**
- 6: **break**
- 7: **end if**
- 8: **if** $k_i > K_i$ **then**
- 9: **continue**
- 10: **end if**
- 11: Select e^* according to Formula 4.
- 12: $ES = ES \cup \{e^*\}; E = E - \{e^*\}$.
- 13: $k_i = k_i + 1$
- 14: **end for**
- 15: **end while**
- 16: **return** ES

In Formula 4, ES represents the expansion terms already selected; $\beta \in [0, 1]$ controls the trade-off between relevance and redundancy of the expansion terms, which is experimentally set to 0.5; $\operatorname{sim}(\vec{e}, \vec{\alpha}_i)$ denotes the similarity between an expansion term and the aspect under consideration; $\operatorname{sim}(\vec{e}, \vec{e}')$ is the similarity between two expansion terms. Both $\operatorname{sim}(\vec{e}, \vec{e}')$ and $\operatorname{sim}(\vec{e}, \vec{\alpha}_i)$ are computed as the dot product of the corresponding vectors, as defined in Formula 5 and 6, respectively:

$$\operatorname{sim}(\vec{e}_i, \vec{e}_j) = \vec{e}_i^T \cdot \vec{e}_j \quad (5)$$

$$\operatorname{sim}(\vec{e}, \vec{\alpha}) = \vec{e}^T \cdot \vec{\alpha} \quad (6)$$

where \vec{e}^T (\vec{e}_i^T) means the transpose of \vec{e} (\vec{e}_i). Note that the above two similarities fall into the scope of $[-1, 1]$, since expansion term vectors and aspect vectors are normalized to 1 (in terms of Frobenius-norm).

Expansion Terms Generation. Once we get ES , a set of diversified expansion terms, we combine them with the original query to form a new query, in which each expansion term has a weight reflecting how relevant it is to the original query:

$$w(e) \leftarrow \sum_{i=1, 2, \dots, N} |\sigma_i| \cdot \vec{e}^T \cdot \vec{\alpha}_i \quad (7)$$

$$w(e) \leftarrow \frac{w(e)}{\sum_{e' \in ES} w(e')} \quad (8)$$

Formula 7 sums up the relevance of the term to all the aspect vectors and also considers the weights of the aspect vectors. The weight of an expansion term w.r.t. the new query is normalized to $[0, 1]$ using Formula 8. Finally, we submit the composed query to Indri, and obtain a set of search results, which is considered as the diversified search results.

Experiments

Data Preparation. We conduct experiments on the ClueWeb09 (category B) dataset, which has 50,220,423 documents (about 1.5 TB), and use the test queries from TREC 2009, 2010 and 2011 Web tracks. We use the log data of Microsoft Live Search 2006 as the resource to generate expansion terms, which spans over one month (starting from May 1st) consisting of almost 14.9M queries shared between around 5.4M user sessions. Since spam filtering is known to be an important component of Web retrieval, we have applied the publicly available Waterloo spam ranking⁴ to the ClueWeb09 (B) collection, and we consider a percentile of 60% which is shown to be optimal for the ClueWeb dataset.

Parameter Setting. We build a development data set consisting of 10 randomly selected queries (from TREC 2009, 2010 and 2011 Web tracks) to fine tune parameters, and obtain the following configuration: β , the trade-off parameter of MMRE defined in Formula 4 is set to 0.5; η , the parameter that controls the trade-off between Frobenius-norm loss and trace norm loss, is set to 1; the number of dimensions of an expansion term vector (N) is fixed to 30; K , the number of expansion term candidates, is set to 100; the parameter that controls the trade-off between relevance and non-redundancy in $MMRE_Q$ (Bouchoucha, Liu, and Nie 2014) is set to 0.6. Notice that the results that we report later for $MMRE_Q$ are calculated on 15 expansion terms (excluding the original query terms). Finally, before running our system, one should initialize each column vector \vec{e} . In our experiments, we set each dimension to $\frac{1}{\sqrt{N}}$ which ensures that $\|\vec{e}\|_F^2 = 1$, which does not promote any aspect to another. We have found that this setting works well in practice and makes our framework converge to a low-rank vector space for the query.

Reference Systems. For comparison purpose, we consider the following reference systems: *BL*, the basic retrieval system that is built with Indri and is based on a query generative language model with Dirichlet smoothing ($\mu=2000$), Krovetz stemmer, and stopwords removal; *MMR*, the system based on the re-ranking of search results (with $\lambda = 0.6$) (Carbonell and Goldstein 1998); *PM-2*, a term-level search result diversification system (Dang and Croft 2012) that considers aspect popularity (Note that, in this work, we didn't implement PM-2 and we simply report the results from the paper of (Dang and Croft 2013)); *xQuAD*, a probabilistic framework for search result diversification, which explicitly models an ambiguous query as a set of sub-queries (Santos, Macdonald, and Ounis 2010); $MMRE_Q$, a MMRE based system, which uses query logs to get diversified expansion terms (Bouchoucha, Liu, and Nie 2014).

Evaluation Metrics. We consider the following official measures as performance metrics: nDCG and ERR (Chapelle et al. 2009) for adhoc relevance performance, α -nDCG (Clarke et al. 2008) (in our experiments, $\alpha=0.5$), ERR-IA (Chapelle et al. 2009), NRBP (Clarke, Kolla, and Vechtomova 2009) and Prec-IA (Agrawal et al. 2009) for diversity measure, and S-recall (Zhai, Cohen, and Lafferty 2003) to measure the ratio of covered subtopics for a given

query.

Results. We report the performance numbers in Table 1 and Table 2 on queries of TREC 2009, 2010, 2011 and their combination, respectively. From Table 1 and Table 2⁵, we observe that our method consistently outperforms all the other systems in terms of both relevance and diversity on all data sets, and the improvements are statistically significant for all the measures. This observation confirms the overall advantage of our proposed method. In particular, in Table 2, comparing our approach with other state-of-the-art approaches on the 144 queries used in the previous experiments (Dang and Croft 2013), we can see that our approach outperforms *PM-2* and *xQuAD* by large margins.

From Table 1, we can see that our method performs significantly better than $MMRE_Q$, a typical DQE approach, on all the measures. This is a clear indication of the advantage to use low-rank aspect vector space to represent the possible intents of a query. We can analyze one particular query "rice" in TREC 2010 Web track⁶, which is a typical example that shows the general trends. This query is ambiguous and has five different subtopics. For this query, $MMRE_Q$ outputs the following expansion terms: *rice, recipe, diet, university, dish, condoleezza, baseball, cooker, sushi, resume, chinese, chicken, pictures, houston, pudding, facts*; in contrast our method yields less expansion terms which however covers more aspects of the query: *rice, recipe, sushi, calories, condoleezza, mexican, nutrition, university, biography, chinese, facts, pudding*. We make several observations here. First, some expansion terms like *dish, pictures, cooker* and *houston* are dropped by our method. Such terms with general meanings do not clearly correspond to any subtopic of the query, or could be related to several aspects. In the aspect vector space, such terms represent concrete instances of the semantic category denoted by the terms, but are not close to either of them. In contrast, other terms like *mexican* and *biography* that are not considered by $MMRE_Q$ were nevertheless selected by our method, because they are more specific terms that clearly correspond to narrow subtopics of the query. Second, we observe that our system does not select some terms such as *baseball* and *pictures* that are less important i.e. they don't match any of the manual (ground-truth) query subtopics. This could be explained by the fact that our system not only selects good expansion terms but also ensures a high coverage of the ground-truth query subtopics. Interestingly, one can see that the fourth subtopic of the query was not covered by $MMRE_Q$ but our system covers it through terms *calories, nutrition* and *facts*. Third, we find that our method successfully keeps several expansion terms like *recipe, condoleezza* and *university* which represent one or several aspects of the query. This ensures a high coverage of the query subtopics despite the fact that some subtopic (subtopic 5) is still missing from the top selected expansion terms. This example clearly shows that our method can efficiently identify a set of expansion terms with less redundancy while covering more aspects of the query, largely ow-

⁵Since we don't have the results of PM-2 for each query, we can not run statistical significance test for PM-2.

⁶<http://trec.nist.gov/data/web/10/wt2010-topics.xml>

⁴<https://plg.uwaterloo.ca/~gvcormac/clueweb09spam>

Queries	Method	nDCG@20	ERR@20	α -nDCG@20	ERR-IA@20	NRBP	Prec-IA@20	S-recall@20
TREC 2009	BL	0.312	0.125	0.297	0.195	0.162	0.111	0.430
	MMR	0.310	0.119	0.296	0.191	0.161	0.120	0.442*
	MMRE _Q	0.346*-	0.138*-	0.345*-	0.210*-	0.190*-	0.136*-	0.489*-
	Ours ($\eta = 0$)	0.384*+ [‡]	0.161*+ [‡]	0.392*+ [‡]	0.255*+ [‡]	0.228*+ [‡]	0.167*+ [‡]	0.585*+ [‡]
	Ours ($\eta = 1$)	0.419*+[‡]	0.175*+[‡]	0.417*+[‡]	0.273*+[‡]	0.246*+[‡]	0.185*+[‡]	0.639*+[‡]
TREC 2010	BL	0.182	0.139	0.320	0.203	0.163	0.170	0.543
	MMR	0.191*	0.142	0.329	0.213*	0.170	0.172	0.562*
	MMRE _Q	0.217*-	0.155*-	0.361*-	0.220*-	0.200*-	0.203*-	0.573*-
	Ours ($\eta = 0$)	0.253*+ [‡]	0.179*+ [‡]	0.418*+ [‡]	0.277*+ [‡]	0.245*+ [‡]	0.256*+ [‡]	0.617*+ [‡]
	Ours ($\eta = 1$)	0.270*+[‡]	0.194*+[‡]	0.442*+[‡]	0.298*+[‡]	0.266*+[‡]	0.270*+[‡]	0.659*+[‡]
TREC 2011	BL	0.298	0.139	0.542	0.440	0.399	0.240	0.764
	MMR	0.304	0.141	0.544	0.433	0.397	0.250	0.741
	MMRE _Q	0.351*-	0.149*-	0.600*-	0.478*-	0.428*-	0.261*-	0.776*-
	Ours ($\eta = 0$)	0.387*+ [‡]	0.166*+ [‡]	0.658*+ [‡]	0.572*+ [‡]	0.488*+ [‡]	0.307*+ [‡]	0.824*+ [‡]
	Ours ($\eta = 1$)	0.396*+[‡]	0.182*+[‡]	0.673*+[‡]	0.594*+[‡]	0.509*+[‡]	0.322*+[‡]	0.851*+[‡]

Table 1: Experimental results of different methods on TREC Web tracks query sets. *, -, +, \S indicate significant improvement ($p < 0.05$ in T-test) over BL, MMR, MMRE_Q, and Ours ($\eta = 0$), respectively.

Queries	Method	nDCG@20	ERR@20	α -nDCG@20	ERR-IA@20	NRBP	Prec-IA@20	S-recall@20
TREC 2009 + 2010 + 2011 (144 queries)	BL	0.267	0.133	0.385	0.279	0.241	0.179	0.578
	MMR	0.263	0.131	0.387	0.278	0.240	0.179	0.579
	PM-2	0.305	0.154	0.464	0.344	0.309	0.206	0.627
	xQuAD	0.305*-	0.152*-	0.437*-	0.314*-	0.278*-	0.207*- \S	0.617*-
	MMRE _Q	0.303*-	0.149*-	0.433*-	0.308*-	0.271*-	0.198*-	0.611*-
	Ours ($\eta = 0$)	0.334*+ [‡]	0.162*+ [‡]	0.481*+ [‡]	0.366*+ [‡]	0.324*+ [‡]	0.239*+ [‡]	0.682*+ [‡]
Ours ($\eta = 1$)	0.359*+[‡]\otimes	0.180*+[‡]\otimes	0.505*+[‡]\otimes	0.379*+[‡]\otimes	0.333*+[‡]\otimes	0.251*+[‡]\otimes	0.724*+[‡]\otimes	

Table 2: Comparison of our method with existing SRD methods. *, -, +, \S , and \otimes indicate significant improvement ($p < 0.05$ in T-test) over BL, MMR, xQuAD, MMRE_Q, and Ours ($\eta = 0$), respectively.

ing to its low-rank vector space representation of the aspects of the query. Finally, the trace norm regularization is one of the key points of this work. To understand how the trace element impacts on our method, we set $\eta = 0$ in Formula 1 and report their results. From Table 1 and Table 2, it is clear to see that the difference between our method when using trace norm ($\eta = 1$) is statistically significant compared to our method when the trace element is ignored ($\eta = 0$). This highlights the important role of the trace norm which successfully selects good expansion terms, which improves search results in terms of both relevance and diversity.

Discussions. In this work, we set N , the number of dimensions of an expansion term vector, to 30, independent of the query under consideration. It is interesting to check if our method is sensitive to the setting of N . To do this, we vary $N = 5, 10, 20, 30, 50, 60$ and compare the performance of our system on the development data set. The results are plotted in Figure 1. From $N = 5$ to 10 to 20, we observe that the performance of our system drastically increases on both relevance and diversity measures. This is because a low value of N does not usually ensure the coverage of all the query aspects; the larger N is, the better our system performs. $N = 30$ corresponds to the best setting which yields the highest scores on all the measures. Interestingly, we find that when N increases ($N = 50$ and $N = 60$), the performance of our system is almost the same as the best achieved when $N = 30$. This suggests that our system is not sensitive to N when it is big enough to cover all the aspects of the query. This is indeed one advantage of our system: even though the value of N increases, our frame-

work converges to a vector space that yields similar expansion terms. On the other hand, if the value of N is small, its performance degrades and even underperforms *BL* (case of $N = 5$) when N becomes significantly smaller than the optimal value. Since a larger N introduces more computations to resolve the optimization problem defined in Formula 1, ideally given a query, we should set N to the corresponding optimal value. However the optimal value is not available in practice. In this work, we deal with this problem by fixing N to an arbitrary number with the hope that it is not less than the maximum optimal value w.r.t. all the queries. Note that this is another advantage of our method over LDA, in which we have to carefully pre-decide the number of topics. We use a low-rank vector space to represent all the aspects of a query. It is still not clear how the manually annotated sub topics can be aligned with the learnt aspect vectors. It is also meaningful to study how to use this representation to directly generate diverse search results without a middle step of generating diverse query expansion terms, e.g., mapping a document into the same vector space and choosing a set of diversified documents by running an algorithm similar to Algorithm 1. In this work, we use one resource to generate expansion term candidates. Several recent studies show that using multiple resources can yield better results for SRD (Bouchoucha, Liu, and Nie 2014; He, Hollink, and de Vries 2012). It is thus interesting to extend our method to support multiple resources in future. Finally, in this paper, we adopted a uniform distribution when initializing each aspect vector ($\frac{1}{\sqrt{N}}$ for each dimension). However, this may not be the best setting for initial

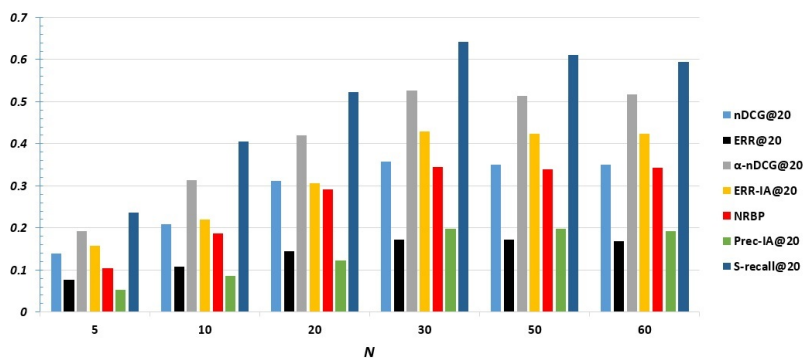


Figure 1: Performance of our system when varying the number of dimensions of an expansion term vector on the development dataset.

values. Therefore, it will be interesting to test several other settings of initial values, in our future work.

Conclusions and Future Work

A DQE approach uses one or several resources to generate a set of diverse expansion terms to obtain a better coverage of the different aspects of a query. Its focus is mainly on removing redundant expansion terms, and it still remains unclear how the expansion terms cover the aspects. In this paper, we propose to use trace norm regularization to learn a low rank vector space to explicitly represent the aspects of a query: we select a special basis consisting of eigenvectors to represent aspects, and the absolute values of the eigenvalues to represent the association strength of the aspect with the query. In the low rank vector space, we greedily select representative expansion terms for each aspect. By doing so, our method ensures that the selected expansion terms not only are different among them, but also can better cover the underlying query aspects. Our experiments on TREC data confirm the effectiveness of our method.

In the future, we want to explore three directions. First, we are interested in mapping search results into the same aspect vector space and performing search results diversification directly in the vector space. Second, we will investigate whether a close correspondence with the manual subtopics is crucial in DQE. Finally, since expansion term candidates from different resources can complement each other, it is desirable to extend our method to support multiple resources.

References

Agrawal, R.; Gollapudi, S.; Halverson, A.; and Ieong, S. 2009. Diversifying search results. In *Proc. of WSDM*, 5–14.

Blei, D. M.; Ng, A.; and Jordan, M. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3:993–1022.

Bouchoucha, A.; He, J.; and Nie, J.-Y. 2013. Diversified query expansion using conceptnet. In *Proc. of CIKM*, 1861–1864.

Bouchoucha, A.; Liu, X.; and Nie, J.-Y. 2014. Integrating multiple resources for diversified query expansion. In *Proc. of ECIR*, 98–103.

Cai, J.-F.; Candès, E. J.; and Shen, Z. 2010. A singular value thresholding algorithm for matrix completion. *SIAM J. on Optimization* 20(4):1956–1982.

Carbonell, J., and Goldstein, J. 1998. The use of mmr, diversity-based reranking for reordering documents and producing summaries. In *Proc. of SIGIR*, 335–336.

Chapelle, O.; Metzler, D.; Zhang, Y.; and Grinspan, P. 2009. Expected reciprocal rank for graded relevance. In *Proc. of CIKM*, 621–630. ACM.

Clarke, C. L.; Kolla, M.; Cormack, G. V.; Vechtomova, O.; Ashkan, A.; Bütcher, S.; and MacKinnon, I. 2008. Novelty and diversity in information retrieval evaluation. In *Proc. of SIGIR*, 659–666.

Clarke, C. L.; Kolla, M.; and Vechtomova, O. 2009. An effectiveness measure for ambiguous and underspecified queries. In *Proc. of ICTIR*, 188–199. Springer-Verlag.

Dang, V., and Croft, W. B. 2012. Diversity by proportionality: An election-based approach to search result diversification. In *Proc. of SIGIR*, 65–74.

Dang, V., and Croft, B. W. 2013. Term level search result diversification. In *Proc. of SIGIR*, 603–612.

Dou, Z.; Hu, S.; Chen, K.; Song, R.; and Wen, J.-R. 2011. Multi-dimensional search result diversification. In *Proc. of WSDM*, 475–484. ACM.

He, J.; Hollink, V.; and de Vries, A. 2012. Combining implicit and explicit topic representations for result diversification. In *Proc. of SIGIR*, 851–860.

He, J.; Meij, E.; and de Rijke, M. 2011. Result diversification based on query-specific cluster ranking. *Journal of the American Society of Information Science and Technology* 62(3):550–571.

Huang, P.-S.; He, X.; Gao, J.; Deng, L.; Acero, A.; and Heck, L. 2013. Learning deep structured semantic models for web search using click-through data. In *Proc. of CIKM*, 2333–2338.

Ji, S., and Ye, J. 2009. An accelerated gradient method for trace norm minimization. In *Proc. of ICML*, 457–464.

Koren, Y.; Bell, R.; and Volinsky, C. 2009. Matrix factorization techniques for recommender systems. *Computer* 42(8):30–37.

Ma, S.; Goldfarb, D.; and Chen, L. 2011. Fixed point and bregman iterative methods for matrix rank minimization. *Math. Program.* 128(1-2):321–353.

Nguyen, D.-P., and Hiemstra, D. D. 2012. Ensemble clustering for result diversification. In *Proc. of TREC*, 83.

Recht, B.; Xu, W.; and Hassibi, B. 2008. Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization. In *CDC*, 3065–3070.

Santos, R. L.; Macdonald, C.; and Ounis, I. 2010. Exploiting query reformulations for web search result diversification. In *Proc. of WWW*, 881–890.

Speer, R., and Havasi, C. 2012. Representing general relational knowledge in conceptnet 5. In *Proc. of LREC*.

Tomioka, R., and Aihara, K. 2007. Classifying matrices with a spectral regularization. In *Proc. of ICML*, 895–902.

Vargas, S.; Santos, R. L. T.; Macdonald, C.; and Ounis, I. 2013. Selecting effective expansion terms for diversity. In *Proc. of OAIR*, 69–76.

Zhai, C. X.; Cohen, W. W.; and Lafferty, J. 2003. Beyond independent relevance: Methods and evaluation metrics for subtopic retrieval. In *Proc. of SIGIR*, 10–17.

Zhu, X.; Goldberg, A.; Gael, J. V.; and Andrzejewski, D. 2007. Improving diversity in ranking using absorbing random walks. *HLT-NAACL* 97–104.