International Journal of New Computer Architectures and their Applications (IJNCAA) 8(2): 65-77
The Society of Digital Information and Wireless Communications, 2018 ISSN 2220-9085 (Online); ISSN 2412-3587 (Print)

Diversifying Search Result Leveraging Aspect-based Query Expansion

Md Shajalal[a*] Masaki Aono[b*] Muhammad Anwarul Azim[c*]
shajalal@baust.edu.bd aono@tut.jp azim@cu.ac.bd
[*]Department of Computer Science and Engineering
[a]Bangladesh Army University of Science and Technology, Nilphamari, Bangladesh
[b]Toyohashi University of Technology, Toyohashi, Aichi, Japan
[c]University of Chittagong, Chittagong, Bangladesh

ABSTRACT

Web search queries are short, ambiguous and tend to have multiple underlying interpretations. To reformulate such queries, query expansion is a prominent method that leads to retrieve a set of relevant documents. In this paper, we propose an aspect-based query expansion technique for diversified document retrieval. At first, query suggestions and completions are retrieved from major commercial search engines. A frequent phrase-based soft clustering algorithm is then applied to group similar retrieved candidates into clusters. Each cluster represents different query aspect. The expansion terms are selected from the generated cluster labels for each cluster. To estimate the relevancy between the expanded query and the documents, multiple new lexical and semantic features are introduced using the content information, and word-embedding model, respectively. Finally, a linear ranking approach is employed to re-rank the documents retrieved for the original query using the extracted features. We conduct experiments on Clueweb09 document collection using TREC 2012 Web Track queries. The experimental results clearly demonstrate that our proposed aspect-based query expansion method is effective to diversify the retrieved documents and outperformed baseline and some known related methods in terms of diversity metrics ERR-IA, α-nDCG and NRBP at the cut of 20.

KEYWORDS

Query Ambiguity, Query Expansion, Diversified Search, Query Aspect, and Word Embedding

1 INTRODUCTION

Web search has become the predominant method for users to fulfill their information needs. In this aspect, users describe their information needs by providing a set of keywords. These keywords are collectively called a search query for each user. Since expressing information need through keywords is difficult, some users fail to choose the precise terms while others tend to omit important terms needed to clarify search intentions [1, 2]. Therefore, a large number of the web search queries are usually short, ambiguous, and prone to have multiple interpretation [3, 4, 5]. Generally, the short queries mean a lot of ambiguity as to what information needs the users express. Consider a short and ambiguous query "*Java*", which could be interpreted as a programming language, island, coffee, etc.

For such type of queries, the search engine may generate a ranking of documents with maximum redundancy covering a very few user information needs. To mitigate these issues, search result diversification (SRD) can be used to generate the effective ranking of documents. Some clustering algorithms applied in different perspectives [6, 7] can be used for SRD. Diversification approaches re-rank the retrieved documents considering intents or aspects for the user query. Therefore, the retrieved documents contain less redundant documents. In turns, the retrieved documents also cover user query aspects as much as possible. The common principle used in the existing SRD approaches is to select as diverse results as possible from a given set of retrieved documents. The final ranking list is much dependent on the initial retrieval results, which may not have a good coverage of the different aspects of the query.

To overcome this drawback, some existing studies on SRD attempted to expand the original query before diversifying the results [8, 9].

Query expansion is a classic technique to reformulate the query, which generates diversified expansion terms to enhance the original query. There are lot of approaches to expand the original query using different resources and techniques including pseudo-relevance feedback [10], word-embedding [11, 12, 13], ConceptNet and WordNet [14], and Freebase [15] etc. Query expansion techniques are widely applied for improving the efficiency of the textual information retrieval systems. These techniques help to overcome vocabulary mismatch issues by expanding the original query with additional relevant terms and reweighing the terms in the expanded query.

In this research, we propose an aspect-based query expansion technique to diversify the retrieved documents for the original query. Query suggestions and completions from search engines are good resources to reformulate the original query. Therefore, our proposed approach is to retrieve query suggestions and completions for each query from three commercial web search engines namely, Google, Yahoo, and Bing. The aggregated list of suggestions and completions are used as a resource to expand the original query. A frequent phrase based soft clustering algorithm is then applied to group similar candidates into clusters. Here every cluster represents different query aspect. The generated cluster labels are then used to expand the query. We employ all the terms from the cluster labels except query terms and stop words to expand the query. Finally, the retrieved documents are re-ranked based on the relevancy with the expanded query. To estimate the relevancy between web document and the query, we propose multiple semantic and lexical features using word-embedding and their content information, respectively. We conducted experiments using the Clueweb09 document collection with TREC 2012 Web Track query set. The experimental results clearly illustrate that our proposed aspect-based query expansion method is effective to diversify web documents. There are two distinct contributions in proposed method:

1. A novel query expansion technique based on users' aspect and

2. Multiple new semantic and lexical features to estimate the relevance between expanded query and documents

The rest of the paper is structured as follows: In section **2**, we summarize related work on query expansion and document retrieval. In section **3**, we briefly explain two classical retrieval model. We present our proposed method in section **4**. The experiments and evaluation to show the effectiveness of our proposed method is presented in section **5**. Some concluded remarks and future directions are described in section **6**.

2 RELATED WORK

Usually, queries to web search engines are short and not written carefully, which makes it more difficult to understand the intent behind a query and retrieve relevant documents. A common solution is query expansion, which uses a larger set of related terms to represent the user's intent and improves the documents' ranking.

Pseudo Relevance Feedback (PRF) algorithms are widely used in query expansion. These algorithm assume that top ranked documents for the original query are relevant that contain good expansion terms. The researchers proposed a model that selected expansion terms based on their term frequency in top retrieved documents, and weights them by documents' ranking scores [16]:

$$s(q) = \sum_{d \in D} p(t|d) f(q, d)$$

where D is the set of top retrieved documents, $p(t|d)$ denotes the probability that term t that generated by document d's language model, and $f(q, d)$ denotes the ranking score of the document provided by the retrieval model. Later, another study [17] added inverse document frequency (IDF) to demote very frequent terms:

$$s(q) = \sum_{d \in D} p(t|d) f(q, d) \log \frac{1}{p(t|C)}$$

where $p(t|C)$ denotes the probability of term t in the corpus language model C.

Another PRF approach has also been proposed using a Mixture Model [18]. In that study, the researchers assumed that the terms in top retrieved documents are drawn from a mixture of two language models: query model θ_q and a background model θ_B. The likelihood of a top retrieved document d is defined as follows:

$$\log p(d|\theta_q, \alpha_d, \theta_B) = \sum_{t \in D} \log(\alpha_d\, p(t|\theta_q) + (1 - \alpha_d)\, p(t|\theta_B))$$

α_d denotes a document-specific mixture parameter. For this equation, the query model θ_q can be learned by maximizing the top retrieved documents' likelihood. The terms that have non-zero probability in θ_q are used for query expansion.

Knowledge base such as Freebase can also be applied by query expansion methods [15]. Those methods identified the entities associated with the query, and used the entities to perform query expansion. A supervised model combined information derived from Freebase descriptions and categories to select terms that are effective for query expansion. The researchers also proposed a method to expand the original query with the help of WordNet and ConceptNet [14]. Their approach extended the query with the synonyms generated from WordNet.

Recently, word-embedding techniques are used for query expansion [11, 12, 13]. An Automatic Query Expansion (AQE) framework has been proposed by using distributed neural language model, word2vec [13]. They trained a word2vec model that learned a low dimensional embedding for each vocabulary entry using the semantic and contextual relation in a distributed and unsupervised approach. They selected the related terms to the query by applying a K-nearest neighbor technique and those terms were used for expansion. A query expansion technique is introduced for *adhoc* retrieval using a locally trained word embedding model [11]. They presented local embedding which capture the nuances of topic-specific language better than global embeddings. Another study also proposed word-embedding based method for query expansion [12]. They applied continuous-bag-of-words implementation of word2vec over the entire corpus on which search is performed and selected terms that are semantically related to the query. Their method either used the terms to expand the original query or integrate them with the effective pseudo-feedback-based relevance model.

3 CLASSICAL RETRIEVAL MODEL

In this section, we discuss two classical retrieval models, Okapi BM25 [19] and Language model [16]. In this research, these two classical retrieval models are utilized as baseline document retrieval.

3.1 Okapi BM25 Model

Let d be an unstructured document in the collection C. We may consider this as a vector $\vec{d} = (tf_1, \cdots, tf_V)$, where tf_i denotes the term frequency of the *i*-th term t_i in the document d and V is the total number of terms in the vocabulary. In order to score such a document against a query, most ranking functions define a term weighting function $w_j(\vec{d}, C)$. BM25 is an example of such functions. For ad-hoc retrieval, and ignoring any repetition of terms in the query, BM25 [19] can be simplified as follows:

$$w_j(\vec{d}, C) = \frac{(k_1 + 1) \cdot tf_j}{k_1(1 - b) + b\frac{dl}{avdl}} \log \frac{N - df_j + 0.5}{df_j + 0.5} \tag{1}$$

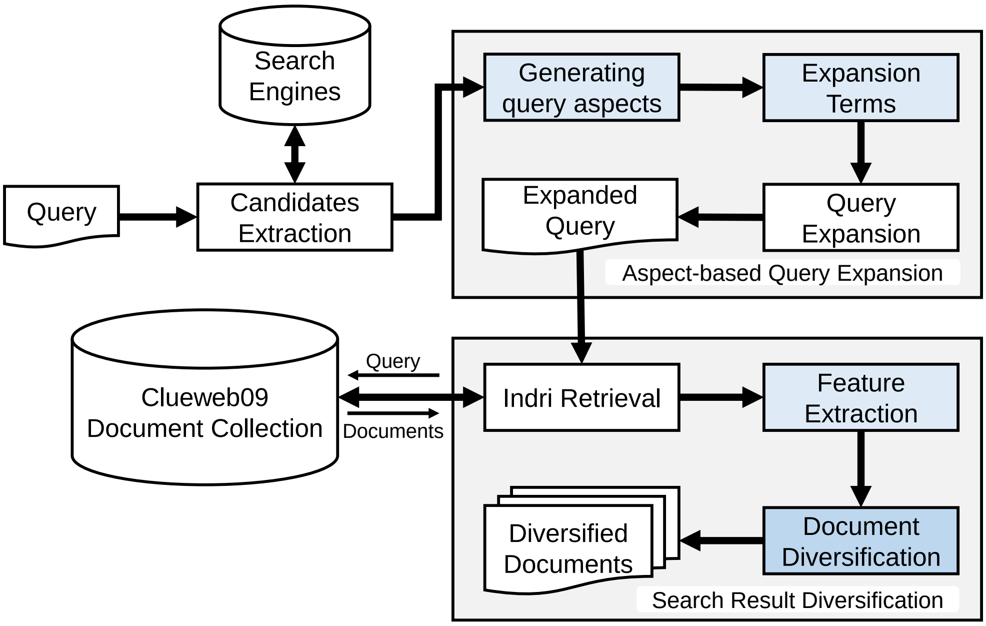

Figure 1. An Aspect-based Query Expansion Framework for SRD

where df_j is the document frequency of *j*-th term, dl is the document length, $avdl$ is the average document length in the collection, and k_1 and b are tuning parameters.
The document score is then obtained by adding the document term weights of term matching the query q:

$$W(\vec{d}, q, C) = \sum_j w_j(\vec{d}, C) \cdot tf_{q_j} \qquad (2)$$

3.2 Language Model

Language model is a quite general formal approach in information retrieval. Query likelihood model is the most basic method for using language models in information retrieval. Let us assume a simple unigram for each document, where each document is represented as the standard bag-of-words and their language model is distributed over a vocabulary of a single word. The maximum likelihood estimate of term w occurring in document d for a multinomial distribution is given below [16]:

$$P_{ML}(q|d) = \frac{tf_{w,d}}{|d|} \qquad (3)$$

where $tf_{w,d}$ is term frequency (number of times term w appears in document d) of the term w in document d and $|d|$ denotes the total number of terms in d.

Given a query $q = \{q_1, q_2, q_3 ... q_k\}$, the likelihood can be computed for the document d as follows:

$$P(q|d) = \prod_{i=1}^{k} P(q_i|d) \qquad (4)$$

This likelihood is computed for each document and used for ranking. Ranking documents in this procedure is known as query likelihood language model.
The smoothed $p(w|d)$ with Jelinek-Mercer smoothing is estimated as follows:

$$p(w|d) = \lambda \frac{tf_{w,d}}{|d|} + (1-\lambda)\frac{cf_w}{|C|} \qquad (5)$$

where cf_w is the term frequency of term w in the collection C and λ is a smoothing parameter.

4 OUR PROPOSED METHOD

This section presents our proposed aspect-based query expansion method for search result diversification. For a given query, our method produces a diversified list of documents using aspect-based query expansion. The high-level building blocks of our proposed method are illustrated in Fig. 1. There are two major parts in our method, *query expansion*

Table 1. Multiple candidates reflect one aspect.

Query	Candidate	Aspect
grilling	memorial day grilling recipes easy grilling recipes grilling recipes	Grilling Recipes
	grilling chicken grilling chicken breasts grilling chicken leg	Grilling Chicken
	grilling corn on the cob grilling corn	Grilling Corn
	grilling lobster tails grilling lobster	Grilling Lobster
	grilling tips outdoor grilling tips perfect grilling tips	Grilling Tips

and *search result diversification* using the expanded query.

Given a query, the *query expansion* technique expands the original query covering user query aspects as much as possible. In this regards, we propose a new aspect-based approach to select expansion terms. The *search result diversification* technique returns a list of diversified documents with respect to the expanded query. In this part, multiple new semantic and lexical features are introduced to estimate the relevancy between the query and document. The remainder of this section presents the complete explanation of the *query expansion* and the *search result diversification* techniques.

4.1 Aspect-based Query Expansion

Generally, search queries are very short in length. The existing study on query structure suggested that, the average length of search queries is around 2.3 terms per query [20]. Usually, the short queries mean a lot of ambiguity as to what information needs the users express. Therefore, we reformulate the original query by appending more related terms that reflect different user aspects. This process is called query expansion. Here our expansion method tries to select terms with various aspects underlying a query as much as possible. Our hypothesis is that, if the query covers more user aspects, the task of estimating the relevancy between the expanded query and documents will be more easier. The original query is expanded using the following three steps.

4.1.1 Candidate Extraction

Query reformulation with related query suggestions and completions are more effective for searching the most relevant documents that maximize the coverage [8, 21]. The query suggestions and completions from commercial search engines are employed as a resource to find the expansion terms. We retrieve query suggestions and query completions from major search engines (Google, Yahoo, and Bing) for a given query. Then the duplicates are filtered out after aggregating all suggestions and completions.

4.1.2 Generating Query Aspects

Multiple query suggestions and completions may contain candidates which reflect the same query aspect. Our observation on this aspect is that a group of candidates covers similar query aspect rather covering unique aspects. Table 1 depicts an example where we can see that multiple candidates represent the same aspect. In this table, we can see five aspect of query "*grilling*" including "*grilling recipes*", "*grilling chicken*", "*grilling corn*", "*grilling*

Table 2. lexical features

Type	Features	Description
Lexical Features	01. $f_{LS}(q_{exp}, d)$	Lexical similarity based on edit distance
	02. $f_{TO}(q_{exp}, d)$	% of overlapping query terms
	03. $f_{SynO}(q_{exp}, d)$	% of overlapping synonym of query terms.
	04. $f_{BR}(q, d)$	Baseline rank of each individual document
	05. $f_{VisTerm(d)}$	Number of visible terms on the page
	06. $f_{TTerm}(title(d))$	Number of terms in the page <title>field
	07. $f_{avgTL}(d)$	Avg. length of visible terms on the document
	08. $f_{fracAT}(d)$	Fraction of anchor text on the document
	09. $f_{fracVT}(d)$	Fraction of visible text on the document
	10. $f_{fracS}(d)$	Stopword and Non-stopword ratio

lobster", and "*grilling tips*".

A soft clustering technique is then applied to the candidates based on frequent phrases to identify the query aspects. We make use of Lingo Clustering algorithm [22] to group the candidates into clusters. Some candidates may belong to more than one cluster. Then we used the cluster labels generated by the clustering algorithm as query aspects.

4.1.3 Selecting Expansion Terms

For a given query q, let assume that $L_q = \{l_1, l_2, l_3, ..., l_K\}$ be the set of cluster labels generated by the previous section. We generate and select the expansion terms from these labels to expand the query. To select the expansion terms from the generated cluster labels, we introduce a expansion term selection algorithm. The pseudo-code of our expansion term selection (ETS) algorithm is as follows:

Algorithm 1: Expansion Term Selection: ETS(q, L_q)

Input: Set of generated cluster labels for query q, $L_q = \{l_1, l_2, l_3, ..., l_K\}$
Output: Set of expansion terms, E_t
$E_t = [\emptyset]$;
for *each term t* $\in L_q$ **do**
  if $t \in l$ && $t \notin q$ **then**
    $E_t = E_t \cup t$
  end
end

where E_t denotes the set of expansion terms, l is the cluster label, and t is the term in label l. $t \in l$ and $t \notin q$ state that term t exists in l and t does not exist in q, respectively.

4.1.4 Query Expansion

Let $E_t = \{t_1, t_2, t_3, ..., t_n\}$ be the set of selected terms for query q. We make use of these terms to expand the query. We append the selected terms with the query q and the expanded query is as like as follows:

$$\begin{aligned} & q_{exp} = q \\ & for\ each\ term\ t;\ t \in E_t \\ & \quad q_{exp} = q_{exp} \cup t \end{aligned} \tag{6}$$

where q_{exp} denotes the expanded query.

4.2 Search Result Diversification

This section presents the diversification approach to re-rank the retrieval result for original query. The documents are re-ranked with respect to their relevancy between the expanded query q_{exp} and the documents. We first extract multiple semantic and lexical features, then we apply a linear ranking function to rank the documents. Since the expanded query covers multiple query aspects, the re-ranked documents satisfy the diversity.

4.2.1 Feature Extraction

The lexical and semantic features are estimated using WordNet [23] and the content information of document and expanded query and the

Table 3. Semantic features

Type	Features	Description
Semantic Features	01. $f_{MPP}(d)$	$\frac{\sum_{t \in d} I(POS(t) \in M)}{\lvert d \rvert}$
	02. $f_{ACS}(q_{exp}, d)$	$\frac{\sum_{t_i \in q_{exp}} \sum_{t_j \in d} ConSim(t_i, t_j)}{\lvert q_{exp} \rvert * \lvert d \rvert}$
	03. $f_{SIM_{w2v}}(q_{exp}, d)$	$\frac{\sum_{t_i \in q_{exp}} \sum_{t_j \in d} cosine(\vec{t_i}, \vec{t_j})}{\lvert q_{exp} \rvert * \lvert d \rvert}$

pre-trained word2vec[1] model on *Google News Corpus*. The lexical and semantic features are summarized in Table 2 and 3, respectively.

The notations in the Table 3 are defined as follows:

- In Meaningful POS (part-of-speech) percentage (MPP) feature $f_{MPP}(d)$, $I(POS(t) \in M)$ returns 1 if the POS of term t belongs to the set $M = \{Noun, Verb, Adjective, Adverd\}$ [24].

- In average concept similarity (ACS) feature $f_{ACS}(q_{exp}, d)$, $ConSim(t_i, t_j)$ returns the conceptual similarity between term t_i and t_j [24].

- In similarity (SIM) feature $f_{SIM_{w2v}}(q_{exp}, d)$ based on *Word2Vec*, $\vec{t_i}$ and $\vec{t_j}$ denote the 300 dimensional vector representation of term t_i and t_j from pre-trained word2vec model, respectively.

We make use of *MinMax* normalization to normalize the features value into the range [0,1] as follows:

$$\bar{x} = \frac{x - min(x)}{max(x) - min(x)}$$

where x is the feature value and $\bar{x}$ is the normalized feature value. $min(x)$ and $max(x)$ denote the minimum and maximum feature values of a specific feature, respectively.

[1] word2vec (https://code.google.com/p/word2vec/)

4.2.2 Document Ranking

To re-rank the retrieved documents for original query, we estimate the document relevancy using a linear ranking approach considering all extracted features as follows:

$$\begin{aligned} Rel(q_{exp}, d) &= \frac{\sum_{i=1}^{N} wt_i \cdot f_i(q_{exp}, d)}{\sum_{i=1}^{N} wt_i} \\ &= \frac{wt_1 \cdot f_1(q_{exp}, d)}{\sum_{i=1}^{N} wt_i} + \frac{wt_2 \cdot f_2(q_{exp}, d)}{\sum_{i=1}^{N} wt_i} \\ &+ \cdots + \frac{wt_N \cdot f_N(q_{exp}, d)}{\sum_{i=1}^{N} wt_i} \end{aligned} \tag{7}$$

where $f_i(q_{exp}, d)$ denotes the *i*-th feature and wt_i denotes the feature importance. The higher the value of $Rel(q_{exp}, d)$, the higher the relevancy of the document d with the expanded query q_{exp} is.

5 EXPERIMENTS AND EVALUATION

This section presents the details of the experiments and evaluation results of our proposed method on a standard dataset and compare with some known related methods.

5.1 Data Collection

We use the Web Track dataset from TREC 2012 [25]. There are 50 queries, each of which includes 3 to 8 subtopics identified by TREC assessors. All experiments are conducted on ClueWeb09 [26] collection. We used query suggestions and completions from Googe, Yahoo, and Bing search engine provided by NTCIR-10 English subtopic mining dataset [27]. The pre-trained Word2Vec model using Google News Corpus are employed to extract semantic features. To find the POS of each term, we make use of Stanford NLP Parser. We make use of Indri search engine [28] to retrieve top 500 documents from the clueweb09 collection.

5.2 Evaluation Metrics

Several metrics have been used in order to evaluate the diversification effectiveness of search engines. A good diversification system is the one that satisfies multiple information needs (or user intents) underlying a query that is submitted to that system by different users, or by the same user in different contexts. In the context of search result diversification, a query is represented by a set of subtopics or aspects (which generally correspond to user intents). The relevance of a document with respect to a query is judged separately for each subtopic, and is estimated by the ability of that document to cover different subtopics of the same query. In this research, we utilized three diversity metrics which are official in the diversity task of TREC Web track.

5.2.1 α-nDCG (α-normalized Discriminative Cumulative Gain)

α-nDCG@k [29] is computed as follows:

$$\alpha - nDCG@k = \frac{\alpha - DCG@k}{\alpha - DCG'@k}$$

where $\alpha - DCG'@k$ is a normalization factor corresponding to the maximal value of $\alpha - DCG@k$ that gives the ideal document ranking. $\alpha - DCG@k$ is computed as follows:

$$\alpha - DCG@k = \sum_{j=1}^{k} \frac{\sum_{s \in S(q)} rel(d_j, s)(1-\alpha)^{\sum_{i=1}^{j-1} rel(d_i,s)}}{log_2(1+j)}$$

where the parameter α ($\alpha \in$ [0,1]) represents the user satisfaction factor for the set of documents that have been already browsed by the user. This parameter (α) is generally fixed to 0.5. q is a query, $S(q)$ is the set of subtopics underlying q, and d_i (resp. d_j) is the document ranked at the ith (resp. jth) position. $rel(d, s)$ is a function that evaluates the relevance of a document d with respect to a given subtopic s. Note also that $\alpha - nDCG$ considers the set of already $(k-1)$ selected documents when evaluating a document at position k. This means that the metric takes into account the dependency between the returned documents. Finally, note that $(1-\alpha)^{\sum_{i=1}^{j-1} rel(d_i,s)}$ penalizes the coverage of already covered aspects of the query and α controls the amount of penalization.

5.2.2 ERR-IA (Expected Reciprocal Rank - Intent Aware)

ERR-IA(q, D) [30] for a given query q and over a set of returned documents D with respect to q is defined as follows:

$$ERR - IA@k = \sum_{s \in S(q)}^{k} p(s|q) \cdot ERR(s, D)$$

where $ERR(s, D)$ is the expected reciprocal rank and $p(s|q)$ denotes the importance of subtopic s regarding to the query q (the more popular the subtopic s for q, the higher is $p(s|q)$.

5.2.3 NRBP (Novelty and Rank-Biased Precision)

NRBP [31] is an extension of the RBP (Rank-Biased Precision) metric [32]. The basic intuition that NRBP uses is that, the user has some specific intent and is generally interested in one particular aspect of the query, at least at that time. NRBP is defined as follows:

$$NRBP = \frac{1-(1-\alpha)^{\beta}}{N} \cdot \sum_{k=1}^{\infty} \beta^{k-1} \cdot \sum_{i=1}^{N} J(d_k, i)(1-\alpha)^{C(k,i)}$$

where d_k denotes the kth document, N is the (possible) number of aspects of a given query, $J(d, i) = 1$ if document d is relevant to the ith aspect of the query , and $J(d, i) = 0$ otherwise, $C(k, i)$ is the number of documents at cut-off k that have been judged to be relevant to the ith aspect of the query, parameter $\beta \in$ [0,1] is used to model the patience level of the user, and parameter $\alpha \in$ [0,1] refers to the user declining interest.

In short, we use the above three official diversity evaluation metrics used in TREC Web Track.

Table 4. Summary of all experimental settings.

Run	Description
LM	Documents retrieved using language model
BM25	Documents retrieved with BM25 retrieval model
QFLR	Original query and linear ranking with all features
Q_{n_exp}FLR	Query expansion with suggestions's terms and linear ranking with all features
Q_{a_exp}LFLR	Aspect-based query expansion and linear ranking with lexical features
Q_{a_exp}FLR	Aspect-based query expansion and linear ranking with lexical and semantic features

Table 5. Experimental results of our method, baseline and some known related methods on TREC Web Track 2012 in terms of ERR-IA, α-nDCG, and NRBP at cut of 20. **Boldface** indicates the **best** performance ammong all.

Type	Method	ERR-IA@20	α-nDCG@20	NRBP
Baseline Retrieval	BM25 [19]	0.2253	0.3105	0.1738
	LM [16]	0.157889	0.2143	0.1237
Our Method	**Q_{a_exp}FLR**	**0.3447**	**0.4438**	**0.3033**
	Q_{a_exp}LFLR	0.3015	0.3925	0.2685
	Q_{n_exp}FLR	0.2475	0.3475	0.2348
	QFLR	0.2354	0.3257	0.1925
Related Methods	ICTNET [25]	0.326	0.422	0.280
	udel [25]	0.325	0.419	0.282
	LIA [25]	0.318	0.424	0.268
	udel fang [25]	0.300	0.420	0.241

5.3 Experimental Results

To measure the effectiveness of our aspect based query expansion method for SRD, we carried out experiments using different experimental settings. We retrieved top 500 documents for each query by using two baseline retrieval model, Language model with Jelinek-Mercer smoothing and Okapi BM25. We already described these to classical model in section 3. We used the parameter *mu*=2500, *lambda*=0.4 for Jelinek-Mercer smoothing and *k1*=1.2, *b*=0.75 and *k3*=7 for BM25 based method. We denote these two experimental settings as LM and BM25, respectively.

We carried out experiments using our reranking method presented on section 4.2 with documents retrieved by BM25. In setting QFLR, we re-ranked the retrieved documents by using a linear ranking method (i.e Eq. 7)with extracted features with respect to the original query. Then we expanded the query with the query suggestions' terms except stopwords and query terms. This setting is denoted by Q_{n_exp}FLR. Then we applied our aspect-based query expansion method to expand the query and linear ranking with only lexical features to re-rank the documents in setting Q_{a_exp}LFLR. Finally, in setting Q_{a_exp}FLR, we used all features instead for lexical features to re-rank the documents with respect to the expanded query q_{exp}. Table 4 summaries the description of all experimental settings.

The performance of our proposed methods, baseline, and some related [25] methods on TREC Web track 2012 dataset in terms of three diversity metrics including ERR-IA, α-nDCG, and NRBP at the cut of 20 are reported in Table 5. Boldface indicates the best performance among all methods. We can see that, our two aspect-based query expansion methods Q_{a_exp}LFLR and Q_{a_exp}FLR performed better

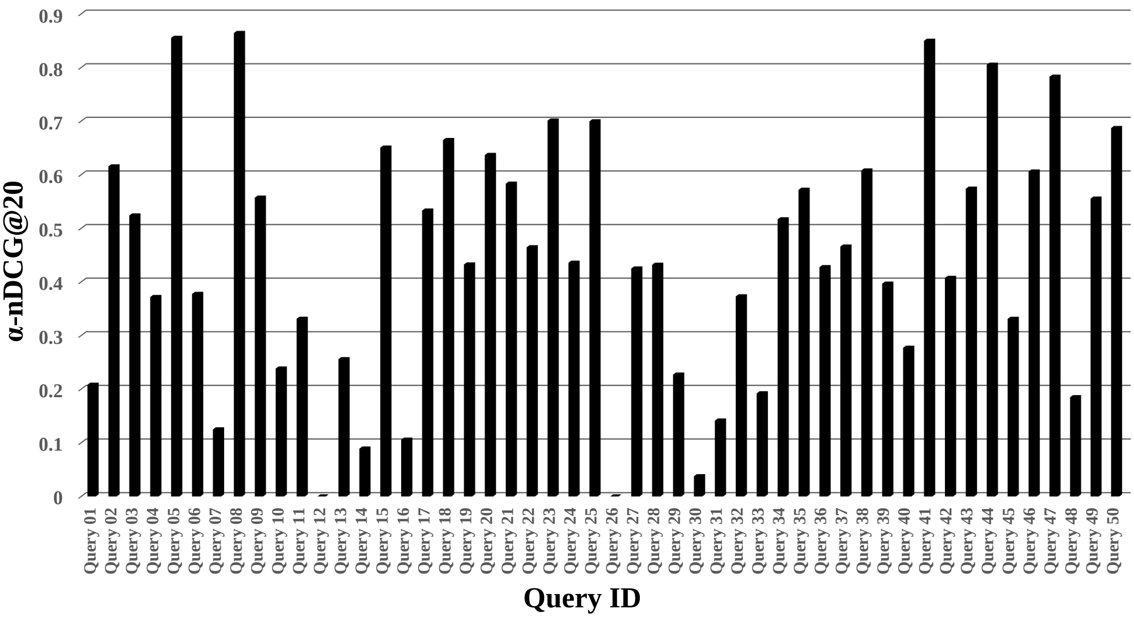

Figure 2. The query-wise performance of our proposed method method in terms of α-nDCG@20. The X-axis represents individual Query-ID and Y-axis represents the value of α-nDCG@20 for each individual query.

than all other methods including the normal query expansion based method Q_{n_exp}FLR. Therefore, we can conclude that aspect-based query expansion can capture different query aspects which helps to increase the diversity of the ranking. Our proposed semantic features are not applied in setting Q_{a_exp}LFLR whereas those are applied in setting Q_{a_exp}FLR. The experimental results clearly demonstrate that our semantic features are effective to capture better relevancy.

The query-wise performance of our method in terms of the diversity metric α-nDCG@20 on TREC Web Track 2012 dataset is depicted in Fig. 2. The figure illustrates that the performance for each individual query are varied widely. We can see that, our method achieved more than 80% accuracy (i.e Query 05, Query 09, Query 41, etc.) for several queries . For an example query “porterville” (Query 09), our method achieved 85% accuracy. The aspect-based query expansion technique selected the expansion terms for this queries are: “recorder,” “college,” “ca,” “school,” “district,” “fair,” “weather,” “police,” “department,” and “courthouse”. These expansion terms are distinct from each other and all are related to the query. We can also see that each term is representing different user’s query aspect for the query “porterville”. The figure also conclude that our method failed for only two queries (Query 12 and Query 26). Considering the query, “dnr” (Query 12) which is an abbreviation query. The observation for this query is that our frequent phrase-based clustering algorithm was not good enough to generate meaningful cluster labels. In turns the expansion terms were not related to the abbreviation queries. We think this might be one of the plausible reasons for the failure. However, we can conclude from the experimental results on a benchmark dataset that our proposed aspect-based query expansion method and semantic features contributed to effectively diversify the documents. Our method also outperformed the baselines and known related methods in terms of all three standard diversity evaluation metrics.

6 CONCLUSION AND FUTURE DIRECTIONS

This paper proposed an aspect-based query expansion method and multiple semantic features for search results diversification. To identify the query aspects, we applied a frequent phrase-based soft clustering technique to the query suggestions. Then we select the expansion terms from the cluster labels. We also proposed multiple semantic and lexical features to estimate the relevancy between the expanded query and the document. The experimental results on a benchmark TREC dataset clearly conclude that our proposed method is effective for search result diversification.

For future work, we have a plan to extract expansion terms from top retrieved web document for query expansion. Furthermore, it will be interesting to apply an aspect-based document diversification approach for results diversification.

REFERENCES

[1] Se-Jong Kim, Jaehun Shin, and Jong-Hyeok Lee. Subtopic mining based on three-level hierarchical search intentions. In *European Conference on Information Retrieval*, pages 741–747. Springer, 2016.

[2] Se-Jong Kim and Jong-Hyeok Lee. Subtopic mining using simple patterns and hierarchical structure of subtopic candidates from web documents. *Information Processing & Management*, 51 (6):773–785, 2015.

[3] Rodrygo LT Santos, Craig Macdonald, Iadh Ounis, et al. Search result diversification. *Foundations and Trends® in Information Retrieval*, 9(1):1–90, 2015.

[4] Kevyn Collins-Thompson, Craig Macdonald, Paul Bennett, Fernando Diaz, and Ellen M Voorhees. Trec 2014 web track overview. Technical report, MICHIGAN UNIV ANN ARBOR, 2015.

[5] Behrooz Mansouri, Mohammad Sadegh Zahedi, Maseud Rahgozar, Farhad Oroumchian, and Ricardo Campos. Learning temporal ambiguity in web search queries. pages 2191–2194, 2017.

[6] Rahmat Widia Sembiring, Jasni Mohamad Zain, and Abdullah Embong. Dimension reduction of health data clustering. *International Journal of New Computer Architectures and their Applications (IJNCAA)*, pages 1018–1026, 2011.

[7] Yongxin Zhang Hong Shen Shitian Xu. Cluster sampling for the demand side management of power big data. *International Journal of New Computer Architectures and their Applications (IJNCAA*, pages 114–121, 2016.

[8] Rodrygo LT Santos, Craig Macdonald, and Iadh Ounis. Exploiting query reformulations for web search result diversification. In *Proceedings of the 19th International Conference on World Wide Web*, pages 881–890. ACM, 2010.

[9] Rodrygo LT Santos, Jie Peng, Craig Macdonald, and Iadh Ounis. Explicit search result diversification through sub-queries. In *European Conference on Information Retrieval*, pages 87–99. Springer, 2010.

[10] Mohannad Almasri, Catherine Berrut, and Jean-Pierre Chevallet. A comparison of deep learning based query expansion with pseudo-relevance feedback and mutual information. In *European conference on information retrieval*, pages 709–715. Springer, 2016.

[11] Fernando Diaz, Bhaskar Mitra, and Nick Craswell. Query expansion with locally-trained word embeddings. *arXiv preprint arXiv:1605.07891*, 2016.

[12] Saar Kuzi, Anna Shtok, and Oren Kurland. Query expansion using word embeddings. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 1929–1932. ACM, 2016.

[13] Dwaipayan Roy, Debjyoti Paul, Mandar Mitra, and Utpal Garain. Using word embeddings for automatic query expansion. *arXiv preprint arXiv:1606.07608*, 2016.

[14] Arbi Bouchoucha, Xiaohua Liu, and Jian-Yun Nie. Towards query level resource weighting for diversified query expansion. In *European Conference on Information Retrieval*, pages 1–12. Springer, 2015.

[15] Chenyan Xiong and Jamie Callan. Query expansion with freebase. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, pages 111–120. ACM, 2015.

[16] Victor Lavrenko and W Bruce Croft. Relevance-based language models. In *ACM SIGIR Forum*, volume 51, pages 260–267. ACM, 2017.

[17] Chengxiang Zhai and John Lafferty. A study of smoothing methods for language models applied to ad hoc information retrieval. In *ACM SIGIR Forum*, volume 51, pages 268–276. ACM, 2017.

[18] Hamed Zamani, Javid Dadashkarimi, Azadeh Shakery, and W Bruce Croft. Pseudo-relevance feedback based on matrix factorization. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pages 1483–1492. ACM, 2016.

[19] Andrew Trotman, Antti Puurula, and Blake Burgess. Improvements to bm25 and language models examined. In *Proceedings of the 2014 Australasian Document Computing Symposium*, page 58. ACM, 2014.

[20] Claudio Carpineto and Giovanni Romano. A survey of automatic query expansion in information retrieval. *ACM Computing Surveys (CSUR)*, 44(1):1, 2012.

[21] Youngho Kim and W Bruce Croft. Diversifying query suggestions based on query documents. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 891–894. ACM, 2014.

[22] Claudio Carpineto, Stanislaw Osiński, Giovanni Romano, and Dawid Weiss. A survey of web clustering engines. *ACM Computing Surveys (CSUR)*, 41(3):17, 2009.

[23] Christiane Fellbaum. *WordNet*. Wiley Online Library, 1998.

[24] Md Shajalal, Md Zia Ullah, Abu Nowshed Chy, and Masaki Aono. Query subtopic diversification based on cluster ranking and semantic features. In *Advanced Informatics: Concepts, Theory And Application (ICAICTA), 2016 International Conference On*, pages 1–6. IEEE, 2016.

[25] Ian Soboroff, Iadh Ounis, Craig Macdonald, and Jimmy J Lin. Overview of the trec-2012 microblog track. In *TREC*, volume 2012, page 20, 2012.

[26] Jamie Callan, Mark Hoy, Changkuk Yoo, and Le Zhao. Clueweb09 data set, 2009.

[27] Tetsuya Sakai, Zhicheng Dou, Takehiro Yamamoto, Yiqun Liu, Min Zhang, Ruihua Song, MP Kato, and M Iwata. Overview of the ntcir-10 intent-2 task. In *NTCIR*, 2013.

[28] Bhaskar Mitra, Fernando Diaz, and Nick Craswell. Luandri: a clean lua interface to the indri search engine. *arXiv preprint arXiv:1702.05042*, 2017.

[29] Charles LA Clarke, Maheedhar Kolla, Gordon V Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. Novelty and diversity in information retrieval evaluation. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 659–666. ACM, 2008.

[30] Olivier Chapelle, Shihao Ji, Ciya Liao, Emre Velipasaoglu, Larry Lai, and Su-Lin Wu. Intent-based diversification of

web search results: metrics and algorithms. *Information Retrieval*, 14(6): 572–592, 2011.

[31] Charles LA Clarke, Maheedhar Kolla, and Olga Vechtomova. An effectiveness measure for ambiguous and underspecified queries. In *Conference on the Theory of Information Retrieval*, pages 188–199. Springer, 2009.

[32] Alistair Moffat and Justin Zobel. Rank-biased precision for measurement of retrieval effectiveness. *ACM Transactions on Information Systems (TOIS)*, 27(1):2, 2008.