Hyperlink-Extended Pseudo Relevance Feedback for Improved Microblog Retrieval

Tarek El-Ganainy
Qatar Computing Research Institute
Qatar Foundation
Doha, Qatar
telganainy@qf.org.qa

Walid Magdy
Qatar Computing Research Institute
Qatar Foundation
Doha, Qatar
wmagdy@qf.org.qa

Ahmed Rafea
Computer Science Department
American University in Cairo
Cairo, Egypt
rafea@aucegypt.edu

ABSTRACT
Microblog retrieval has received much attention in recent years due to the wide spread of social microblogging platforms such as Twitter. Many research studies investigated different approaches for microblog retrieval. Query expansion is one of the approaches that showed stable performance for improving microblog retrieval effectiveness. In this paper we comprehensively investigate query expansion for microblog retrieval using pseudo relevance feedback (PRF) with different configurations. In addition, we propose a hyperlink-extended PRF by expanding queries using embedded hyperlinks in retrieved microblogs. Our experimental results on TREC microblog data showed that PRF alone could outperform many retrieval approaches if configured properly. Results also showed that utilizing hyperlinked documents in tweets for query expansion improves results over traditional PRF. Our best results achieved using hyperlink-extended PRF outperformed most of the reported results using other approaches.

Categories and Subject Descriptors
H.3.3 [**INFORMATION STORAGE AND RETRIEVAL**]: Query formulation, Relevance feedback, Search process

General Terms
Experimentation, Performance

Keywords
Microblog retrieval, Twitter, Query expansion, Pseudo relevance feedback, Hyperlinked documents

1. INTRODUCTION
Interest in microblog retrieval witnessed significant growth in recent years due to the increased popularity of social microblogging platforms, such as Twitter. The special nature of microblogs, which are short documents (with a maximum of 140 characters), introduced different search scenarios and retrieval challenges when compared to other search tasks, such as web search [[15], [19]]. The major challenge in microblog retrieval lies in the severe vocabulary mismatch between query and short relevant documents. Two main approaches presented effective solutions to this problem, namely query expansion [[3], [5], [8], [10], [12], [13], [14]] and document expansion [[6], [9], [16], [17]]. Most of the reported works on query expansion mainly applied pseudo relevance feedback (PRF) by assuming top retrieved documents relevant and selecting a set of terms to be added to the query for a second search aimed at better vocabulary matching. The numbers of feedback terms and documents in the PRF process were typically selected subjectively in previous studies, leading to the absence of a recommended PRF configuration for a better microblog retrieval performance. In addition, reported work focused on extracting terms from top retrieved documents without considering hyperlinks embedded in many of the retrieved tweets. This represents additional unused content of microblog documents [[5], [8], [10], [12], [13]].

Most of works reported that utilized hyperlinks was to improve retrieval effectiveness by applying document expansion or using them as features for learning to rank algorithms [[6], [11], [16], [17]].

In this paper, we introduce a hyperlink-extended PRF for microblogs by utilizing hyperlinks in tweets in the PRF process to improve retrieval effectiveness. Initially, we extensively experiment with PRF using different configurations to identify to what extent it can improve retrieval effectiveness for microblog retrieval. Later, we apply our hyperlink-extended PRF by extracting additional expansion terms from hyperlinked documents in the feedback tweets. We apply our study on TREC microblog track datasets from 2011, 2012, and 2013.

Comprehensive results for PRF showed that using a smaller number of feedback documents with small weight added to original query achieves much improved results over baseline. Results for hyperlink-extended PRF showed that utilizing embedded hyperlinks in feedback tweets can further achieve significant improvement over traditional PRF. Our best result using hyperlink-extended PRF outperformed systems that used other techniques such as document expansion and learning-to-rank, which motivates the integration of our expansion method with these techniques for further improved results.

The contributions of this paper can be listed as follows:

1. Providing a comprehensive study for the usage of traditional PRF with microblog retrieval using various configurations.
2. Introducing a hyperlink-based PRF approach for microblog retrieval by utilizing hyperlinks embedded in initially retrieved tweets, which showed a significant improvement to retrieval effectiveness.

2. RELATED WORK
TREC microblog track was introduced in 2011 due to the increased interest in microblog retrieval. Ad-hoc search task for

SoMeRA'14, July 11, 2014, Gold Coast, Queensland, Australia.

http://dx.doi.org/10.1145/2632188.2632204

microblogs were studied over the past three years using two tweets collections and three query sets [[16], [17], [9]]. Different approaches were investigated for microblog retrieval [[16], [17], [9]] to overcome the special nature of microblog documents [[15], [19]]. One of the main challenges in microblog retrieval is term mismatch between queries and short relevant documents. Efron et al. [[6]] compared a news article dataset represented in TREC 8 to a tweet dataset represented in TREC 2011 microblog data. They showed that distribution of query terms among the top retrieved search results is strongly peaked due to the short length of tweets, and this negatively affects the retrieval performance. They introduced a document expansion approach to enrich tweets with additional terms by submitting the documents as pseudo queries [[6]]. Their approach managed to improve retrieval effectiveness significantly. Other approaches for document expansion were reported in TREC microblog track that mostly showed improvements to the retrieval effectiveness [[16], [17]]. One of the most applied approaches for document expansion was expanding tweets containing hyperlinks with corresponding titles of the hyperlinked documents. Although document expansion potentially lead to improved retrieval effectiveness for microblog search, its computational cost is high, since using each tweet to search the tweet collection as in [5] or accessing embedded hyperlinks to extract page titles [[16], [17]] for all tweets in a collection is seen impractical for the current large tweets streams. On the other hand, for query expansion only the active search documents that are of interest to users are processed.

An alternative to overcome vocabulary mismatch is query expansion [[2]]. Several studies showed the effectiveness of using query expansion to microblog retrieval [[3], [5], [8], [10], [12], [13], [14]]. Traditional PRF, which selects some terms from initially top retrieved documents, was reported by many participants in TREC microblog track to improve retrieval effectiveness [[16], [17], [9]]. The reported work for PRF selected specific numbers for documents and terms heuristically for the feedback process without a comprehensive study to find the best.

Some work utilized web search as an external resource for microblog query expansion [[1], [4], [5]]. Google API was used to retrieve web results matching microblog query at the same time period of the collection, then terms were selected in a PRF process from the top web results. [[1], [4], [5]] that showed this technique leads to significant improvement over traditional PRF in TREC 2011 and 2013 microblog dataset respectively.

A line of research focused on applying temporal-based query expansion for microblog retrieval [[3], [10], [12], [14]]. Choi et al. [[3]] proposed selecting time period based on user behavior (e.g. retweets) to extract relevant tweets to be used for the expansion process. Massoudi et al. [[10]] showed that selecting terms temporally closer to the query time are more effective in expansion. Similarly, Metzler et al. [[12]] showed that using temporal co-occurrence of terms is much more effective than using the traditional term co-occurrence for choosing expansion terms. In a more recent work, Miyanishi et al. [[14]] proposed a two-stage relevance feedback model using manual tweet selection and incorporating lexical and temporal evidence to the model. They showed that manually selecting one relevant tweet can significantly improve retrieval effectiveness.

In addition to using embedded hyperlinks in tweets for document expansion, hyperlinks have been used widely as features in learning to rank algorithms for improving microblog retrieval [[11], [16], [17]]. McCreadie et al. [[11]] investigated how content of embedded hyperlinks in tweets can help estimating the tweet's relevance. They showed that using hyperlinked documents in tweets improve retrieval effectiveness either by document expansion or through learning to rank algorithm.

Aforementioned work proposed various approaches to improve the retrieval effectiveness for microblog search. Although these approaches are highly advanced and usually led to improvements, they are either computationally costly (e.g. document expansion), dependent on third-party uncontrolled components (e.g. using web-search-based expansion from Google), require user's manual intervention, or require the presence of training data (e.g. learning to rank). Less attention was directed toward the utilization of tweets embedded hyperlinks in query expansion, which is seen to be more efficient and less complicated than other approaches, especially if optimized correctly.

3. PROPOSED APPROACH

A hyperlink in a tweet is more than a link to related content as in webpages, but actually it is considered a link to the main focus of the tweet. In fact, sometimes tweet's text itself is totally irrelevant, and the main content lies in the embedded hyperlink, e.g. "*This is really amazing, you have to check htwins.net/scale2*".

Analyzing the TREC microblog dataset over the past three years, we found more than 70% of relevant tweets contain hyperlinks. This motivates utilizing the hyperlinked documents content in an efficient way for query expansion.

3.1 Deriving Expansion Content

The content of hyperlinked documents in the initial set of top retrieved tweets is extracted and integrated into the PRF process. Titles of hyperlinked pages usually act like heading of the document's content, which can enrich the vocabulary in the PRF process. Similarly, meta-description of a webpage usually summarizes its content, which can be useful to further enrich the expansion terms.

We apply hyperlinked documents content extraction on three different levels:

- Tweets level (**PRF**): which represents the traditional PRF, where terms are extracted from the initial set of retrieved tweets while neglecting embedded hyperlinks.
- Hyperlinked document titles level (**HPRF-1**): where the page titles of the hyperlinked documents in feedback tweets are extracted and integrated to tweets for term extraction in the PRF process.
- Hyperlinked documents meta-description level (**HPRF-2**): the titles and meta-description of hyperlinked documents are extracted and integrated to tweets for term extraction.

Titles and meta-description of hyperlinked documents may include unneeded text. For example, titles usually contain delimiters like '–' or '|' before/after page domain name, e.g., "... | CNN.com" and "... – YouTube". We clean these fields through the following steps:

- Split page titles on delimiters and discard the shorter substring, which is assumed to be the domain name.
- Detect error page titles, such as "404, page not found!" and consider them broken hyperlinks.
- Remove special characters, URLs, and snippet of HTML/JavaScript/CSS codes.

This process helps in discarding terms that are potentially harmful if used in query expansion.

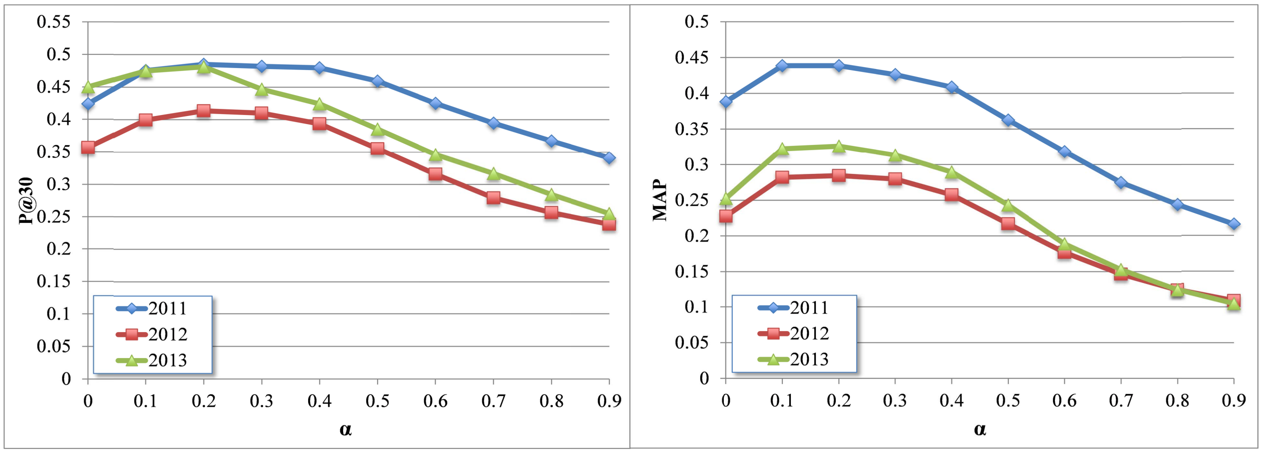

Figure 1. PRF performance for different weight of expansion terms α measured by P@30 and MAP. α ranged from 0 (no expansion) to 0.9.

3.2 Expansion Term Selection

TFIDF and Okapi weighting [[18]] were used for ranking the top terms to be used for query expansion. We calculate TFIDF for a term x as follows:

$$TFIDF(x) = [tf_t(x) + \partial_1 tf_{h_t}(x) + \partial_2 tf_{h_d}(x)] \cdot log\frac{N}{df(x)} \quad (1)$$

Where $tf_t(x)$ is the term frequency of term x in the top n_d initially retrieved tweet documents used in the PRF process; $tf_{h_t}(x)$ is the term frequency of term x in the titles of hyperlinks in the top n_d tweets; and $tf_{h_d}(x)$ is the term frequency of term x in the meta-description of hyperlinks in the top n_d tweets. ∂_1 and ∂_2 are binary functions that equal to 0 or 1 according to the content level of hyperlinked documents used in the expansion process. $df(x)$ is document frequency of term x in the collection; and N is the total number of documents in the collection.

k_1 and b free parameters of the Okapi weighting were selected as 2 and 0 respectively. The parameter b was set to 0 since the variation in tweets length is limited due to Twitter constraint on the number of characters used (max. 140 characters).

Terms extracted from the top n_d initially retrieved documents are ranked according to equation 1, and top n_t terms with the highest TFIDF are used to formulate Q_E for the expansion process. Weighted geometrical mean is used to calculate the final score of retrieval for a given Q query according to equation 2:

$$P(Q|d) = \sqrt{P(Q_0|d)^{1-\alpha} \cdot P(Q_E|d)^{\alpha}} \quad (2)$$

Where Q_0 is the original query; Q_E is the set of extracted expansion terms; $P(Q|d)$ is the probability of query Q to be relevant to document d. α is the weight given to expansion terms compared to original query (when α =0, no expansion is applied). Language-model-based retrieval model was used to calculate the probability of relevance [[7]].

4. EXPERIMENTAL SETUP

4.1 Test Collection

We used the TREC microblog track datasets from 2011, 2012, and 2013 in our experimentation. The datasets include two tweet collections and three query sets. The TREC 2011 microblog collection contains 16 million tweets representing a sample stream between Jan 23 and Feb 7, 2011. Two sets of 50 and 60 topics were released in 2011 and 2012 respectively against this collection along with relevance judgments. A much larger collection of 243m tweets was released in 2013 as a sample stream between Feb. 1 and March 31, 2013 along with 60 topics and relevance judgments.

Table 1. Number of extracted hyperlinks and the number of extracted content of them

dataset	*hyperlinks*	*extracted titles/desc.*	*broken%*
2011	3164	2116 / 1612	33.0%
2012	3781	2360 / 1760	37.6%
2013	3694	2920 / 2099	20.9%

Table 2. Baseline results for TREC microblog test sets

	2011	**2012**	**2013**
P@30	0.4238	0.3565	0.4500
MAP	0.3882	0.2275	0.2524

The two tweets collections were stored and indexed on a server maintained by TREC. Some preprocessing was applied on the corpus such as: tokenization, normalization and stemming. TREC released a set of search APIs to access the corpus. The baseline retrieval offered by the API was based on negative KL-divergence language modeling approach with Dirichlet prior smoothing parameter μ set to 2,500. The baseline retrieval model was based on the Lucene open-source search engine.

We applied a baseline run by searching tweets collections with the original queries Q_0. All hyperlinks appearing in the top 100 results for each query were extracted. Title and meta-description of the corresponding hyperlinked document (webpage) were extracted. We found that on average 60-65% of the results contained embedded hyperlinks, which confirms the importance of utilizing these hyperlinks effectively. Table 1 presents the number of extracted hyperlinks from the top ranked 100 results for each query of each year, and shows the number of extracted page titles and meta-description for these hyperlinks. As shown in Table 1, the percentage of broken hyperlinks for 2011 and 2012 were much larger than 2013. This is expected because of the age of the 2011 tweet collection.

Table 5. Best runs achieved by each query expansion method.
*** and + indicate statistical significant improvement over baseline and PRF**

	P@30			MAP		
	2011	**2012**	**2013**	**2011**	**2012**	**2013**
Baseline	0.4238	0.3565	0.4500	0.3882	0.2275	0.2524
PRF	0.4939*	0.4147*	0.5178*	0.4452*	0.2925*	0.3421*
HPRF-1 (TFIDF)	0.4980*	0.4249*	0.5483*	0.4558*	0.2994*	0.3498*
HPRF-1 (Okapi)	0.4946*	0.4237*	0.5394*+	0.4510*	0.2974*	0.3492*
HPRF-2 (TFIDF)	0.4959*	0.4311*	0.5544*+	0.4575*	0.3034*	0.3570*+
HPRF-2 (Okapi)	**0.5000***	**0.4339***	**0.5546***+	**0.4587***	**0.3044***	**0.3584***+

4.2 Baseline and PRF Runs

Table 2 presents the baseline runs of searching the collections without applying query expansion (i.e. $\alpha = 0$ in equation 2). This baseline is just for comparison with examined PRF methods.

Initially, we applied PRF while setting n_d to 50 and n_t to 10 based on the most reported setup in TREC microblog track [[16], [17]] to find the optimal expansion weight α. We tested several values of α; ranging from 0.1 to 0.9, and monitored the retrieval effectiveness measured by P@30 and MAP, see Figure 1. As shown in Figure 1, $\alpha = 0.2$ achieved the best results for both P@30 and MAP. This is different than literature that typically used 1:1 weighting (α=0.5) between original and expanded query terms. Based on this result, we set $\alpha = 0.2$ for the rest of our experiments.

Once we had α set, we ran different configurations for PRF, HPRF-1, and HPRF-2 using different values of expansion document and terms. n_d values tested were {10, 20, 50, 100}, and n_t values tested were {5, 10, 15, 20, 30}. We aim from this extensive experimentation to study a recommended setup for PRF in microblog retrieval rather than heuristically selecting numbers that may be suboptimal. For all runs, stop-words removal and Porter stemmer were applied.

For the different configurations, we compared the performance of applying hyperlink-extended PRF to traditional PRF via comparing P@30 and MAP, and applying statistical significance test using paired t-test with p-value < 0.05.

5. RESULTS

5.1 Traditional PRF Results

Table 3 reports full results of PRF using tweets content only with different number of feedback documents (n_d) and terms (n_t), while underlining runs that achieved the best scores. Almost all the configurations led to significant improvements over baseline when compared by P@30 and MAP, which confirms the effectiveness of PRF in microblog retrieval. Test sets 2011 and 2013 showed the best performance achieved when using only 10 documents in the feedback process, while 2012 best performance was when using 50 documents in the feedback process. Best results achieved using PRF for TREC 2011 dataset outperformed reported results achieved by other approaches on the same test collection, which was studied extensively in the literature, including: document expansion [[6]], temporal-based query expansion [[3], [13]], web-search-based query expansion [[1]], and learning-to-rank [[13], [16]]. The main reason for this result may be that previous studies did not give much attention to using small number of feedback document in the PRF process, which we showed to achieve the best results for 2011 and 2013 test sets.

5.2 Hyperlink-extended PRF Results

Extending text used for PRF by hyperlinked documents content led to improvements in the retrieval effectiveness. We noticed that HPRF-1 led to less significant improvements than HPRF-2, which indicates that using the meta-description of hyperlinked webpage documents in PRF further improve the results. Tables 4.a-d report full results of different configurations of HPRF-1 and HPRF-2 when using TFIDF and Okapi for term selection. Tables highlight the runs that achieved statistically significant improvement over the corresponding configuration in PRF. Comparing results in Tables 4.a-d to that in Table 3, over 90% of the different configurations of PRF were improved using HPRF. In addition, significant improvements were achieved for many of the PRF configurations of TREC 2013 test set, and some of 2012 test set. In the tables, HPRF-2 achieved better results than HPRF-1 for both TFIDF and Okapi term selection. It was also noticed that Okapi term selection led to slightly better than TFIDF for most of the configurations. The noticed better performance for 2013 test set over 2012 and 2011 may be a result of the percentage of broken hyperlinks of 2012 and 2011 was much higher than 2013 leading to less content to be added for query expansion, see Table 1.

Table 5 summarizes the best runs achieved by each query expansion approach for each test set measured by P@30 and MAP, and compares it to baseline and traditional PRF. As shown, the more content used in PRF process, the better the results. HPRF-1 achieved improvements over PRF, but was not statistically significant, while HPRF-2 achieved further improvements, which made results significantly better than the best runs of PRF for 2013 test set for both P@30 and MAP when using TFIDF or Okapi for term selection. In fact, the best achieved result in Table 5 for 2011 and 2013 test sets outperformed the best reported automatic run in microblog track, which applied query expansion and results re-ranking based on large number of features [[9], [16], [20]].

Our results for hyperlink-extended PRF show that configuring PRF properly and utilizing content of hyperlinked-document in tweets effectively produce an efficient retrieval system that outperforms other sophisticated techniques such as learning-to-rank, where it needs human intervention and costs money to produce training data and document expansion, where it needs to process the whole tweets corpus compared to our proposed method that process only tweets that are of interest to the users. This shows the importance of properly tuning and extending query expansion. Nonetheless, our extended HPRF can be combined with other methods to achieve even better results.

Table 3. PRF results for different number of feedback document and terms for TREC microblog dataset when α=0.2

	n_t	2011					2012					2013				
	n_d	5	10	15	20	30	5	10	15	20	30	5	10	15	20	30
P@30	10	0.4878	**0.4939**	0.4837	0.4830	0.4898	0.3672	0.3763	0.3977	0.3972	0.3977	0.4983	0.5083	0.5156	0.5067	**0.5178**
	20	0.4796	0.4850	0.4918	0.4789	0.4769	0.3870	0.3927	0.3932	0.3977	0.4113	0.4906	0.4972	0.5028	0.4928	0.4922
	50	0.4769	0.4810	0.4850	0.4837	0.4741	0.3881	0.4102	**0.4147**	0.4130	0.4136	0.4944	0.4706	0.4789	0.4822	0.4850
	100	0.4701	0.4769	0.4789	0.4816	0.4810	0.3938	0.3847	0.3972	0.4006	0.4073	0.4883	0.4944	0.4878	0.4756	0.4689
MAP	10	0.4387	**0.4452**	0.4386	0.4398	0.4401	0.2632	0.2651	0.2720	0.2710	0.2721	0.3165	0.3259	0.3374	0.3368	**0.3421**
	20	0.4363	0.4411	0.4382	0.4347	0.4314	0.2678	0.2729	0.2765	0.2796	0.2825	0.3226	0.3291	0.3325	0.3235	0.3254
	50	0.4353	0.4363	0.4387	0.4359	0.4341	0.2748	0.2826	0.2848	0.2855	**0.2925**	0.3229	0.3147	0.3251	0.3233	0.3236
	100	0.4210	0.4267	0.4318	0.4311	0.4276	0.2725	0.2746	0.2803	0.2842	0.2842	0.3153	0.3226	0.3186	0.3141	0.3094

Table 4.a. HPRF-1 (TFIDF) results for different number of feedback document and terms for TREC microblog dataset when α=0.2. Gray cells indicates statistically significant improvement over corresponding PRF configuration

	n_t	2011					2012					2013				
	n_d	5	10	15	20	30	5	10	15	20	30	5	10	15	20	30
P@30	10	0.4864	0.4946	0.4952	**0.4980**	0.4966	0.3842	0.3915	0.3972	0.4073	0.4130	0.5339	**0.5483**	0.5378	0.5267	0.5283
	20	0.4823	0.4905	0.4905	0.4871	0.4816	0.3887	0.3994	0.4164	0.4237	0.4215	0.5101	0.5310	0.5261	0.5194	0.5050
	50	0.4694	0.4796	0.4762	0.4810	0.4728	0.4085	0.4192	**0.4249**	0.4220	0.4226	0.4933	0.5001	0.4972	0.4950	0.4933
	100	0.4782	0.4755	0.4769	0.4741	0.4687	0.4017	0.4068	0.4113	0.4102	0.4001	0.4978	0.4817	0.4733	0.4767	0.4906
MAP	10	0.4381	**0.4558**	0.4525	0.4525	0.4504	0.2703	0.274	0.2781	0.2795	0.2822	0.3314	0.3434	0.3491	0.3471	**0.3498**
	20	0.4376	0.4458	0.4499	0.4405	0.4284	0.2747	0.2836	0.2898	0.2908	0.2937	0.3279	0.3428	0.3450	0.344	0.3433
	50	0.4279	0.4340	0.4305	0.4373	0.4252	0.2827	0.2949	0.2962	**0.2994**	0.2987	0.3264	0.3277	0.3316	0.3358	0.3359
	100	0.4284	0.4278	0.4210	0.4198	0.4164	0.2783	0.2859	0.2901	0.2924	0.2912	0.3159	0.3179	0.3193	0.3188	0.3255

Table 4.b. HPRF-2 (TFIDF) results for different number of feedback document and terms for TREC microblog dataset when α=0.2. Gray cells indicates statistically significant improvement over corresponding PRF configuration

	n_t	2011					2012					2013				
	n_d	5	10	15	20	30	5	10	15	20	30	5	10	15	20	30
P@30	10	0.4741	0.4932	0.4918	0.4946	**0.4959**	0.3836	0.3944	0.4045	0.4169	0.4164	0.5444	0.5489	**0.5544**	0.5433	0.5317
	20	0.4735	0.4884	0.4925	0.4884	0.4918	0.3893	0.3966	0.4028	0.4028	0.4254	0.51	0.5356	0.5411	0.5372	0.5322
	50	0.4741	0.4789	0.4796	0.4789	0.4755	0.4056	0.4056	0.4175	0.4271	**0.4311**	0.505	0.5033	0.5094	0.4967	0.5022
	100	0.4755	0.4714	0.4728	0.4707	0.4687	0.4107	0.4028	0.4113	0.4085	0.4203	0.4761	0.4944	0.485	0.4828	0.4744
MAP	10	0.4385	0.4461	0.4524	**0.4575**	0.4524	0.2705	0.2819	0.2832	0.2842	0.2856	0.3368	0.3503	0.3555	0.3534	**0.357**
	20	0.4365	0.447	0.4455	0.4415	0.4384	0.2783	0.2815	0.2862	0.291	0.297	0.3296	0.3478	0.3546	0.3543	0.3542
	50	0.4288	0.4353	0.4308	0.4318	0.4265	0.2829	0.292	0.2962	**0.3034**	0.3016	0.3299	0.3313	0.339	0.336	0.3399
	100	0.4341	0.4324	0.4321	0.4225	0.4196	0.2802	0.2806	0.2905	0.2907	0.2913	0.3209	0.334	0.3322	0.3293	0.3212

Table 4.c. HPRF-1 (Okapi) results for different number of feedback document and terms for TREC microblog dataset when α=0.2. Gray cells indicates statistically significant improvement over corresponding PRF configuration

	n_t	2011					2012					2013				
	n_d	5	10	15	20	30	5	10	15	20	30	5	10	15	20	30
P@30	10	0.4701	**0.4946**	0.4912	0.4932	0.4945	0.3825	0.3859	0.3910	0.4068	0.4107	0.5344	**0.5394**	0.5339	0.5328	0.5306
	20	0.4687	0.4823	0.4850	0.4884	0.4844	0.3915	0.4006	0.4011	0.4119	0.4198	0.5167	0.5294	0.5139	0.5083	0.5133
	50	0.4714	0.4680	0.4646	0.4714	0.4633	0.4000	0.4147	0.4203	0.4209	**0.4237**	0.4894	0.4956	0.4961	0.4939	0.4922
	100	0.4626	0.4599	0.4599	0.4551	0.4599	0.3989	0.3989	0.4045	0.4079	0.4062	0.4817	0.4711	0.4828	0.4817	0.4717
MAP	10	0.4318	**0.4510**	0.4487	0.4495	0.4493	0.2656	0.2718	0.2739	0.2796	0.2829	0.3306	0.3396	0.3457	0.3452	**0.3492**
	20	0.4245	0.4345	0.4419	0.4349	0.4279	0.2730	0.2819	0.2827	0.2882	0.2941	0.3262	0.3356	0.3355	0.3379	0.3441
	50	0.4251	0.4290	0.4298	0.4285	0.4214	0.2776	0.2848	0.2931	0.2962	**0.2974**	0.3109	0.3197	0.3250	0.3262	0.3285
	100	0.4021	0.4157	0.4135	0.4155	0.4123	0.2704	0.2789	0.2831	0.2865	0.2865	0.3127	0.3072	0.3165	0.3129	0.3123

Table 4.d. HPRF-2 (Okapi) results for different number of feedback document and terms for TREC microblog dataset when α=0.2. Gray cells indicates statistically significant improvement over corresponding PRF configuration

	n_t	2011					2012					2013				
	n_d	5	10	15	20	30	5	10	15	20	30	5	10	15	20	30
P@30	10	0.4680	0.4912	**0.5000**	0.4993	0.4980	0.3842	0.3938	0.4045	0.4136	0.4175	0.5467	**0.5546**	0.5456	0.5417	0.5367
	20	0.4558	0.4823	0.4850	0.4844	0.4871	0.3893	0.3994	0.4096	0.4153	0.4209	0.5050	0.5389	0.5317	0.5289	0.5267
	50	0.4605	0.4707	0.4721	0.4673	0.4680	0.4062	0.4090	0.4107	0.4277	**0.4339**	0.4956	0.4878	0.4939	0.4978	0.5000
	100	0.4592	0.4524	0.4599	0.4592	0.4592	0.3904	0.4034	0.4090	0.4085	0.4051	0.4672	0.4728	0.4700	0.4806	0.4794
MAP	10	0.4310	0.4393	**0.4587**	0.4566	0.4497	0.2655	0.2780	0.2814	0.2832	0.2864	0.3350	0.3421	0.3511	0.3544	**0.3584**
	20	0.4188	0.4391	0.4403	0.4383	0.4367	0.2734	0.2830	0.2898	0.2931	0.2948	0.3189	0.3417	0.3454	0.3464	0.3519
	50	0.4116	0.4246	0.4276	0.4325	0.4205	0.2794	0.2842	0.2908	0.2982	**0.3044**	0.3134	0.3151	0.3207	0.3283	0.3333
	100	0.4009	0.4106	0.4125	0.4147	0.4114	0.2672	0.2752	0.2812	0.2813	0.2891	0.3065	0.3065	0.3073	0.3117	0.3169

6. CONCLUSION AND FUTURE WORK

In this study we investigated the effectiveness of traditional PRF with microblog retrieval, and we proposed a hyperlink-extended version of it that utilizes the presence of embedded hyperlinks in microblogs. Our experimental results on the three datasets of TREC microblog track confirmed the effectiveness of PRF with various kinds of configurations. Results suggest that the optimal weighting between original query and expansion terms is 4:1 (α=0.2) not as the typical weighting 1:1 used in literature. We found that using less number of feedback documents (n_d=10) in the PRF process was more effective than using larger numbers for two of our test sets. It was shown that an effective configuration to PRF parameters could lead to superior results over baseline and outperform other effective approaches. Results for our introduced HPRF method showed that extending feedback documents with additional content from hyperlinked-documents leads to improved results. We found that including the titles and meta-description of hyperlinked-documents to the feedback documents can lead to significant improvement over traditional PRF. Our achieved results (using query expansion only) outperformed most of the reported results for the same datasets, which motivates potential further improvements when other retrieval techniques get integrated. This represents our direct future work. In addition, we plan to investigate the failure point of the expansion technique to work on avoiding it and improve the HPRF method.

7. REFERENCES

[1] A. Bandyopadhyay, K. Ghosh, P. Majumder, and M. Mitra. (2012). Query expansion for microblog retrieval. *International Journal of Web Science.*

[2] C. Buckley, G. Salton, and J. Allan. (1994). The effect of adding relevance information in a relevance feedback environment. *SIGIR 1994.*

[3] J. Choi and W. B. Croft. Temporal models for microblogs. *CIKM 2012.*

[4] A. S. El-Din and W. Magdy. Web-based Pseudo Relevance Feedback for Microblog Retrieval. *TREC 2012*

[5] T. El-Ganainy, Z. Wei, W. Magdy, and W. Gao. (2013). QCRI at TREC 2013 Microblog Track. *TREC 2013*

[6] M. Efron, P. Organisciak, and K. Fenlon. (2012). Improving retrieval of short texts through document expansion. *SIGIR 2012.*

[7] V. Lavrenko and W. B. Croft. (2001). Relevance based language models. In *SIGIR 2001.*

[8] F. Liang, R. Qiang, and J. Yang. (2012). Exploiting real-time information retrieval in the microblogosphere. *JCDL 2012*

[9] J. Lin and M. Efron. (2013). Overview of the TREC2013 microblog track. *TREC 2013.*

[10] K. Massoudi, M. Tsagkias, M. Rijke, and W. Weerkamp. (2011). Incorporating query expansion and quality indicators in searching microblog posts. *AIR 2011.*

[11] R. McCreadie and C. Macdonald. (2013) Relevance in microblogs: enhancing tweet retrieval using hyperlinked documents. *OAIR 2013.*

[12] D. Metzler, C. Cai, and E. Hovy. (2012). Structured event retrieval over microblog archives. *NAACL HLT 2012.*

[13] T. Miyanishi, K. Seki, and K. Uehara. (2013). Combining recency and topic-dependent temporal variation for microblog search. *AIR 2013.*

[14] T. Miyanishi, K. Seki, and K. Uehara. (2013). Improving pseudo-relevance feedback via tweet selection. *CIKM 2013.*

[15] N. Naveed, T. Gottron, J. Kunegis, and A. C. Alhadi. (2011). Searching microblogs: coping with sparsity and document quality. *CIKM 2011.*

[16] I. Ounis, C. Macdonald, J. Lin, and I. Soboroff. (2011). Overview of the trec-2011 microblog track. *TREC 2011.*

[17] I. Soboroff, I. Ounis, C. Macdonald, and J. Lin. (2012). Overview of the TREC-2012 microblog track. *TREC 2012.*

[18] S. E. Robertson, S. Walker, S. Jones, M. Hancock-Beaulieu, and M. Gatford. (1994). Okapi at TREC-3. In *TREC-3*

[19] J. Teevan, D. Ramage, and M. R. Morris. (2012). #TwitterSearch: a comparison of microblog search and web search. *WSDM* 2011.

[20] S. Zhu, Z. Gao, Y. Yuan, H. Wang, and G. Chen. (2013). PRIS at TREC 2013 microblog track. *TREC 2013.*