In Enterprise Search: Methods to identify argumentative discussions and to find topical experts

Maheedhar Kolla
University of Waterloo
Canada. N2L 3G1
mkolla@uwaterloo.ca

Olga Vechtomova
University of Waterloo
Canada. N2L 3G1
ovechtom@uwaterloo.ca

ABSTRACT

Our intention in taking part in this year's Enterprise Track is to experiment with various re-ranking approaches in solving two problems: email search and expert search. In email discussion search, we propose methods to retrieve email messages that contain pro/con argument about the topic in discussion. We compute the likelihood of an email having a pro/con argument about the topic and re-rank emails based on the likelihood of containing topical subjective opinion. In expert search, we explored methods to identify experts for a given topic by analyzing the mailing patterns in the email archive.

1. INTRODUCTION

In this paper, we describe the experiments carried out for discussion search and expert search tasks of Enterprise Search track. In discussion search, we investigated the presence of subjective adjectives towards identifying discussions with pro/con argument about the topic. In expert search, we investigated methods to determine candidate's expertise by analyzing the patterns in the mail archives of the organization. We experimentally compared the performance of graph-based ranking method with another approach using Mutual Information. These experiments were performed using Wumpus [1] search engine and the topics created for this year's Enterprise Search track.

2. DISCUSSION SEARCH

Users interested in knowing more aspects of a certain discussion that might have taken place over a period of time would compose a topic-related query to search the mailing lists. A typical motive behind such search requests is to be able to identify the emails that have some argument—either a pro or a con—about the topic that the user is interested in. A relevant email message should satisfy two constraints:

- It should be on the topic that the user is interested in
- It should have a pro/con argument about the topic in discussion.

In any discussion, participants contribute to its flow by expressing their viewpoints/opinions on the ideas proposed previously (or a new one if the author is initiating the discussion). This is followed by other participants taking turns in either agreeing with the previous author(s) or by suggesting alternative opinions. This motivates our hypothesis that identifying the subjective opinion of the author about the topic, we may be able to retrieve the discussions having a pro/con argument in their content.

Skomorowski [6] proposed a method to determine if some text is expressing opinion targeted towards a topic, by taking into account the adjectives surrounding the topic terms in text. Adjectives, in noun phrases or in the context of nouns, act as noun modifiers. Therefore, by identifying the likelihood of a subjective adjective modifying a noun, he proposed to infer the subjective opinion directed at that noun. He developed a method of resolving adjective targets in a syntactically parsed corpus, i.e. determining which noun is the target of an adjective used either predicatively or attributively. Using a parsed training corpus (AQUAINT), his adjective target resolution method and a set of 1336 subjective adjectives manually composed by Hatzivassiloglou and McKeown [4] he calculated probabilities that at various distances (+/- 10 words) a noun is the target of a subjective adjective. Probability of a noun at distance i being modified by a subjective adjective is calculated as the number of nouns that are targets of subjective adjectives at distance i out of the total number of nouns located at this distance from subjective adjectives. At search time, knowing the distance between the occurrence of the query term and a subjective adjective in a document, the probability of a subjective adjective modifying a noun at that distance precomputed on a training corpus is looked up. The document score is calculated as the sum of such probabilities according to the rule of adding probabilities for non-mutually exclusive events, Inclusion-Exclusion principle (Eq 1)

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1} P(A_i) - \sum_{i<j} P(A_i A_j) + \cdots + (-1)^{n+1} P(A_1 \ldots A_n) \tag{1}$$

where A_i is co-occurrence of a query term with a subjective adjective at distance i in the document; $P(A_i)$ is precomputed probability that at distance i a subjective adjective modifies a noun. We decided to evaluate the effectiveness of the above method in ranking email messages by the likelihood that they contain a pro/con argument towards the topic of the query. We used the probabilities of subjective adjectives modifying nouns at various distances calculated in [6], the document scoring method proposed in [6] and the same list of 1336 subjective adjectives manually composed by Hatzivassiloglou and McKeown [4] in determining the likelihood that an email message contains subjective content targeted at the query, and therefore whether it contains a pro/con argument regarding this topic. Once all email scores are computed, we re-rank the emails based on the updated scores.

3. USING SELECTIVE SUBJECTIVE ADJECTIVES

In the method described in the previous section, all 1336 subjective adjectives were used in re-ranking process. We hypothesize that the frequency of subjective adjectives used in text vary with the subject or topic of the discussion. For example, subjective adjectives such as "incompatible", "unsupported" would be more frequent in in discussions about "hardware devices" and may also express the author's subjective attitude towards the topic of those discussions. So, by selecting only a subset of subjective adjectives, based on the topic, we should be able to retrieve messages that are on topic and have an argumentative style.

To test this hypothesis, we selected only a subset of subjective adjectives using feedback term selection method as proposed by Carpineto et al. [3]. We also experimented with using these selected terms for query expansion.

3.1 Subjective Adjective Selection using KL-Divergence

Carpineto et al. [3] proposed a probabilistic method to select feedback terms based on their distribution in the pseudo-relevant documents (R) and in the documents in the corpus (C). Using the terms from the pseudo-relevant documents, they computed the KL-divergence score for each term as follows:

$$score(t) = [p_R(t)] * log[p_R(t)/p_C(t)] \tag{2}$$

where $p_R(t)$ is the probability of the term t in the relevant (pseudo-relevant) document set; $p_C(t)$ is the probability of the term t in the whole collection. The probability values were computed as follows:

$$p_{(C|R)}(t) = \frac{f_{(C|R)}(t)}{NT_{(C|R)}} \tag{3}$$

where, $f_{(C|R)}(t)$ is the count of the terms in the relevant document set (R), and $NT_{(C|R)}$ is the number of tokens in the document set.

By computing the Kullback-Leibler divergence (KLD) measure as mentioned in the equations above, they were able to determine the terms that maximize the relative entropy—terms that have made the documents in set R divergent to the rest of the documents in set C—as feedback terms. In our method, we used the same 1336 subjective adjectives and compute the KLD scores of the adjectives in the top 25 baseline retrieved documents. We then select the top 40 adjectives from the list of subjective adjectives ranked by their KLD values. Some of the sample feedback terms selected for an example query **"blind HTML authoring tools"** are *accessible, appropriate, disabled, incapable, friendly.*

Using the selected 40 subjective adjectives, we calculated the scores of each email using the BM25 term weighting function [7]:

$$w_t Doc_k = \frac{TF_t(k_1+1)}{k_1 * ((1-b) + b * \frac{|DL|}{AVDL}) + TF_i} * w_t \tag{4}$$

where w_t is the IDF value of the term t; TF_t is the frequency of the term t in document k; DL is the length of the document k and $AVDL$ is the average document length in the collection, k_1 and b are constant values set to 1.2 and 0.5 respectively.

The IDF Value of a term t is computed as follows:

$$w_t = log\frac{(Number\ of\ documents)}{(Number\ of\ documents\ containing\ t)} \tag{5}$$

The sum of the BM25 weights of selected adjective terms in a given document is computed and the sum is used to update the document BM25 score.

3.2 Pseudo-relevance Feedback with subjective adjectives

In addition to using the selected subjectives to re-rank the emails, we used the top 25 subjective adjectives as feedback terms for query expansion. Each subjective adjective term selected was given a weight varying from 0.01 to 2 and the system's performance was evaluated using 2005 Discussion Search topics. We observed that for values 0.03 and 0.04 there was an improvement over the baseline performance. We selected 0.03 for our current experiments. These terms along with the title terms were then used for retrieval process.

3.3 Discussion Search Evaluation

In this year's evaluation, 50 topics were given to the participating systems, which were required to retrieve discussion emails that have a pro/con argument in them. We evaluated the following system submissions:

- uwdsTBaseline : Title only baseline run
- uwdsTrerankSubj : Re-ranking using the subjective adjectives (complete list).
- uwdsTrerankSubjKld : Re-ranking the emails using subjective adjectives selected by based on KL-Divergence values.
- uwdsTExp : Using subjective adjectives for query expansion.

NIST carried out two scale evaluation:

- Performance of system in retrieving messages that are on-topic, but may or may not contain pro/con argument about the topic [Table 1].
- Performance of system in retrieving messages that have pro/con argument about the topic [Table 2][1].

4. EXPERT SEARCH

Expert Search, as defined in this year's Enterprise Search track, is modeled based on the necessity that in an enterprise, users would sometime like to contact a person who possess considerable knowledge about certain topic(s). Along with retrieving the list of candidates for a given topic, it is required to gather a set of evidence documents that could ascertain the affinity of the candidate with the topic. Through

[1]only 46 queries had at least one email message with a pro/con judgement, and so 4 were left out in this evaluation

Table 1: Performance of systems w.r.t retrieval of on-topic emails.

Run	MAP	P@5	P@10	bpref
UWdsTBaseline	0.2898	0.4640	0.4580	0.3084
UWdsTrerankSubj	0.2597	0.4320	0.4280	0.2887
UWdsTrerankSubjKld	0.2493	0.5640	0.4580	0.3501
UWdsTExp	0.3025	0.5280	0.4860	0.3176

Table 2: Performance of systems w.r.t retrieval of emails having pro/con argument only.

Run	MAP	P@5	P@10	bpref
UWdsTBaseline	0.1853	0.2739	0.2522	0.2109
UWdsTrerankSubj	0.1625	0.2174	0.2174	0.2569
UWdsTrerankSubjKld	0.2012	0.3913	0.3283	0.3449
UWdsTExp	0.2048	0.3304	0.2978	0.2941

our experiments we compared the performance of expert retrieval using graph-based ranking approaches like HITS algorithm [5] with other methods such as using out-degree of a candidate and method to rank candidates using Mutual Information.

Like Campbell [2], we first query the W3C mail archive, i.e. "lists" sub-corpus of the W3C collection, and retrieve a set (1000) of email messages. We then extract the candidate names and emails from the header data of each mail, and group them under "From" or "To" groups for each candidate email. Each candidate is assigned two values, *out-degree* - number of emails sent by this candidate in the given set and *in-degree* - number of emails received by the candidate in the set. Using this candidate set, we extract the candidate experts based on the following measures:

4.1 Rank by Outdegree (uwXSOUT)

As a baseline approach for identifying experts for a given query, we rank the candidates based on the number of emails sent by them , Out-degree, in the emails retrieved for a given query. This is based on the idea that candidates who converse frequently about certain topic over emails tend to have knowledge about that topic.

4.2 Graph based ranking (uwXSHUBS)

Kleinberg [5] created a directed graph model of the web, with web pages as nodes and an edge between two nodes signifying the presence of an explicit link. Each node in the graph was assigned with two scores, *hub score* and *authority score*. Kleinberg then proposed an iterative algorithm where in each iteration, the hub scores and authority scores of each node i are updated by looking at the hub and authority scores of the nodes pointing to and being pointed from this node (i).

Inspired by this method, we composed a directed candidate graph and then applied the iterative algorithm as proposed by Kleinberg [5] in order to compute the hub and authority score of each candidate node in the graph. Since we are interested in authoritative candidates, we sorted the candidates based on their authoritative scores.

Table 3: Expert Search: performance without regard to support documents

Run	MAP	P@5	P@10	bpref
uwXSOUT	0.3132	0.5796	0.5143	0.3364
uwXSHUBS	0.2959	0.6000	0.5082	0.3215
uwXSPMI	0.2119	0.3673	0.3653	0.2526

Table 4: Expert Search: performance where candidates retrieved with no positive support documents are considered to be irrelevant

Run	MAP	P@5	P@10	bpref
uwXSOUT	0.1271	0.3306	0.2714	0.1905
uwXSHUBS	0.1389	0.3551	0.2878	0.1938
uwXSPMI	0.0911	0.2245	0.2020	0.1570

4.3 Rank by MI (uwXSPMI)

We compared the graph based ranking approaches with the run in which the candidate expertise is rated based on Mutual Information (MI) score between the candidate name and the topic words. MI score is calculated as follows:

$$MI_{cand,topic} = log\frac{(P_{cand,topic})}{P_{cand}.P_{topic}} \tag{6}$$

Where, P_{cand} is the probability of the candidate in the selected corpus, P_{topic} is the probability of the query term and $P_{cand,topic}$ is the combined probability of candidate and topic co-occurring in the selected corpus. In our experiments, we computed the probabilities using the WWW subcorpus of the W3C collection. Once we compute the MI score between the candidate and query terms, we re-rank the candidates based on the computed scores.

4.4 Expert Search Evaluation

In this year's run of Expert Search task, 55 topics were developed by the participating groups. Each group then judged the topics that they composed. For each topic all the supporting documents (evidence) are judged first and given that there is a positive supporting evidence retrieved for the candidate, the candidate could be judged as an potential expert. In some cases, the judges were allowed to jump to conclusion about the candidate's expertise if they are convinced of the candidate's knowledge by going through only smaller set of evidence.

Two evaluation methods were carried out in this task [2]: 1) Performance measures of the systems taking by evaluating only the candidates retrieved without taking the supporting documents into account. [Table 3] 2) Performance of the systems in which candidate is considered relevant only if the system retrieves at least one positively judged supporting document for the candidate. [Table 4]

5. DISCUSSION AND FUTURE WORK

[2] There were only 49 judged topics

In discussion search runs, methods using only selective subjective adjectives - both re-ranking and query expansion - resulted in better system performance. We also conducted additional experiments in which we used the terms from description for query process. It did not improve the system's performance and so were not included in this paper. We also wish to explore the possibility of combining the subjectivity property with other mailing list properties such as *link structure* that groups emails under same topic into one *thread*.

In the expert search runs, we have observed that the run uwXSOUT has marginally better performance than the other approaches. Using the MI score to re-rank the candidates resulted in poor performance when compared to other runs. We would like to experiment with rank aggregation techniques, by which we could combine both scores from the list and WWW sub-corpus to estimate the candidate's expertise.

6. REFERENCES

[1] S. Buettcher and C. L. Clarke. A security model for full-text search in multi-user environments. In *Proceedings of 4th USENIX Conference on File and Storage Technologies*, 2005.

[2] C. S. Campbell, P. P. Maglio, A. Cozzi, and B. Dom. Expertise identification using email communications. In *CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management*, pages 528–531, New York, NY, USA, 2003. ACM Press.

[3] C. Carpineto, R. de Mori, G. Romano, and B. Bigi. An information-theoretic approach to automatic query expansion. *ACM Trans. Inf. Syst.*, 19(1):1–27, 2001.

[4] V. Hatzivassiloglou and K. R. McKeown. Predicting the semantic orientation of adjectives. In *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, pages 174–181, Morristown, NJ, USA, 1997. Association for Computational Linguistics.

[5] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5):604–632, 1999.

[6] J. Skomorowski. Topical opinion retrieval. Master's thesis, University of Waterloo, (in progress).

[7] K. Sparck-Jones, S. Walker, and S. E. Robertson. A probabilistic model of information retrieval: development and comparative experiments - part 2. *Information Processing and Management*, 36(6):809–840, 2000.