Using Subjective Adjectives in Opinion Retrieval from Blogs

Olga Vechtomova
University of Waterloo
Waterloo, Canada
ovechtom@uwaterloo.ca

1 Introduction

The University of Waterloo participated in the opinion finding task of the Blog track. The task consists of finding blog posts (documents) containing an opinion expressed about the subject of the query. Each query represents a single "entity" that can be, for example, a person, product name, abstract concept, location or event. The relevance judgements were done on a 5-point scale: -1 – not judged, 0 – non-relevant, 1 – relevant, 2 – relevant, negative opinion; 3 – relevant, mixed positive and negative opinion, 4 – relevant, positive opinion.

While many elements in the language can be used to express subjective contents, adjectives are one of the major means of expressing value judgement in English. Our approach consists in using a list of 1336 subjective adjectives manually constructed by Hatzivassiloglou and McKeown [2] for identifying opinionated contents. We hypothesise that presence of subjective adjectives within fixed-size windows around query term instances in a document is a useful feature for finding opinions directed at the query concept.

2 System description

Our approach to finding blogs containing opinions about the concept expressed in the query is a two-staged process. In the first stage, a set of documents is retrieved in response to the query using a topic-relevance ranking method, while in the second stage, this document set is re-ranked using opinion-finding method. The two stages are described in detail below:

In the first stage (document retrieval), 1000 documents are retrieved using BM25 [3] implemented in the Wumpus IR system [1]. The BM25 tuning parameters *b* and *k* were set to 0.75 and 1.2 respectively as they showed optimal performance in [3]. We experimented with two types of queries:

- Single non-stopword terms (runs UWbaseTerms, UWopinion1 and UWopinion2);
- Phrases, which were simply defined as strings of text enclosed in quotation marks by the author of the query (runs UWbasePhrase, UWopinion3 and UWopinion4).

As a result, two baseline runs UWbaseTerms and UWbasePhrase are produced using single terms and phrases respectively.

In the second stage (opinion-finding), we re-ranked the documents retrieved in the previous stage as follows:

For each query term instance in a document, we record all subjective adjectives that occur within the span of *n* words before and after it. We used a list of 1336 subjective adjectives manually composed by Hatzivassiloglou and McKeown [2]. Proximity of the query term instance to each adjective in the window surrounding it is used to calculate the *weighted tf* (*wf*) (Eq. 1 and 2). Stemming was not used in this stage.

We compared the use of stemming with no stemming on the training data from Blog 2006, and better results were obtained without stemming.

$$c(t_i) = \begin{cases} \sum_{a=1}^{A} 1 + \frac{1}{distance(t_i, a)^p} & \text{if } A > 0 \\ 1 & \text{otherwise.} \end{cases} \quad (1)$$

Where: $c(t_i)$ – contribution of the i^{th} instance of the query term t occurring in the document to wf; $distance(t_i,a_j)$ – distance in number of non-stop words between the i^{th} instance of the query term t and subjective adjective a; p - constant, moderating the effect of the distance between t_i and a; A – number of subjective adjectives within the span S before and after t_i. The following span sizes were evaluated: 10, 20 and 30, with 30 giving the best performance on the Blog 2006 data set.

$$wf_t = \sum_{i=1}^{N} c(t_i) \quad (2)$$

Where: N is the number of instances of query term t in the document. After wf is calculated for a query term, its term weight in the document is calculated in the same way as in the BM25 formula, with wf used instead of tf.

After wf is calculated for a query term, its term weight in the document is calculated in the same way as in the BM25 formula [3], with wf used instead of tf (Eq. 3 and 4):

$$TW_t = \frac{(k_1 + 1) \times wf_t}{k_1 \times NF + wf_t} \times idf_t \quad (3)$$

Where: k_1 is the term frequency normalisation factor, which moderates the contribution of the weight of frequent terms. If k_1=0, wf has no effect on the term weight, while the higher the value of k_1 the more effect wf has on the term weight. The value of k_1=1.2 demonstrated optimal results on TREC data sets [3] and was therefore used in all our runs. NF is the document length normalisation factor, and is calculated in the same way as in the BM25 document ranking function, as expressed in Eq. 4.

$$NF = (1 - b) + b \times \frac{DocLen}{AveDocLen} \quad (4)$$

Where: b is a tuning constant, set to 0.75, which showed the best performance in [3], $DocLen$ is the document length in word counts; $AveDocLen$ is the average document length in the document collection. The document score is calculated as the sum of weights of all query terms occurring in it.

3 Evaluation

We compared the use of phrases with the use of single terms in the subjectivity-based document re-ranking algorithm described above, using as the original document set 1000 documents retrieved with single terms (UWbaseTerms run) or phrases (UWbasePhrase run). Table 1 summarises the submitted runs. The results are presented in Table 2.

As can be seen from Table 2, the best opinion relevance results are obtained when the initial document set is retrieved using phrases, while the subjectivity-based re-ranking is done using single terms (run UWopinion3). This run improved performance of 35 topics in MAP compared to the baseline

UWbasePhrase, and of 29 topics in P@10. Figure 1 shows the differences by topic in average precision of the best opinion-finding run UWopinion3 from the baseline UWbasePhrase.

Run name	**Query used to retrieve the original document set (stage 1)**	**Query used in subjectivity-based re-ranking (stage 2)**
UWbaseTerms	single terms	-
UWbasePhrase	phrases	-
UWopinion1	single terms	single terms
UWopinion2	single terms	phrases
UWopinion3	phrases	single terms
UWopinion4	phrases	phrases

Table 1. Summary of the submitted runs

Run name	**Topic relevance**			**Opinion relevance**		
	MAP	**R-Prec**	**P@10**	**MAP**	**R-Prec**	**P@10**
UWbaseTerms (baseline)	0.3231	0.3840	0.6340	0.2426	0.3045	0.4160
UWopinion1	0.3120	0.3674	0.6240	0.2385	0.3083	0.4540
UWopinion2	0.3285	0.3858	0.6480	0.2452	0.3171	0.4460
UWbasePhrase (baseline)	0.3330	0.3936	0.6600	0.2486	0.3087	0.4320
UWopinion3	0.3490	0.4040	0.6800	0.2631	0.3344	0.4960
UWopinion4	0.3384	0.3958	0.6740	0.2511	0.3215	0.4620

Table 2. Evaluation results of the submitted runs

The comparison of four experimental opinion-finding runs shows that using phrases in the first stage (initial document set retrieval) leads to better results than using single words (cf. UWbasePhrase and UWbaseTerms). However, following the phrase-based initial document set retrieval, it is better to use single terms in the second stage (document re-ranking using subjective adjectives). The use of phrases is important in the first stage as it helps to avoid false coordination, however, in the second stage phrases are too restrictive. For example, using a query “brand manager” (topic 929) as a phrase in the first stage is important to avoid documents with “brand” and “manager” used in unrelated contexts (cf. average precision of 0.0386 in UWbaseTerms vs. 0.2077 in UWbasePhrase). However, by using “brand manager” as a phrase in the second stage is less effective than single terms, as the author, for example, having said that “X is a brand manager” may then use the word “manager” throughout the rest of the document (cf. average precision of 0.2298 in UWopinion3 vs. 0.2077 in UWopinion4).

In determining whether to treat a query as a phrase or single terms, we relied on quotation marks placed around phrases by the user. This is not reliable, as not all users enclose phrase queries in quotes. A more reliable automatic method to determine whether a query refers to a single multi-word unit might lead to further improvements.

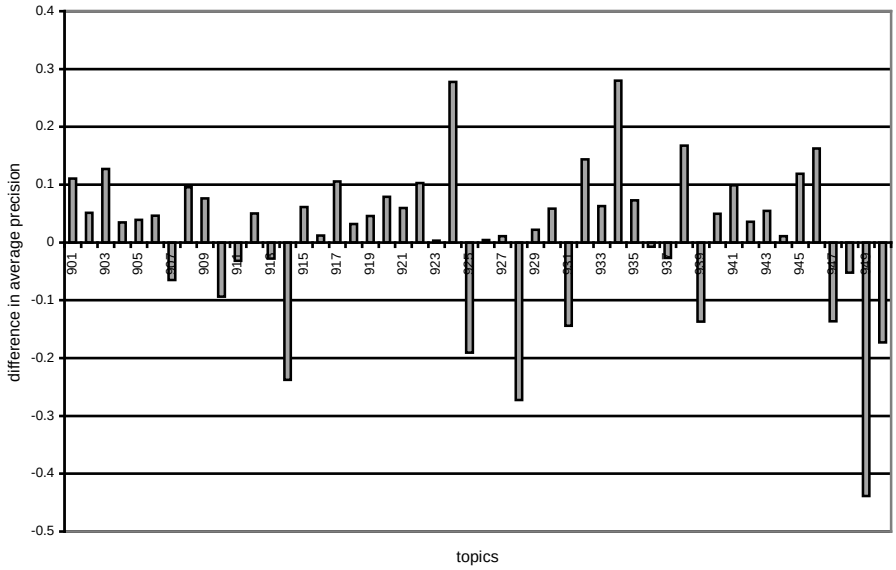

Figure 1. Difference in average precision between UWopinion3 and UWbasePhrase (baseline).

4 Conclusions and Future Work

In this paper we presented a novel method for ranking documents containing opinions about the concept expressed in the query. The method uses a manually constructed list of subjective adjectives, and calculates document score based on the proximity of subjective adjectives to query term instances within fixed-size windows around them in the document. We experimented with different types of queries constructed from the topic titles: single terms and user-defined phrases, i.e. phrases enclosed in quotation marks by the user. Substantial improvements over the baseline (run UWbasePhrase) were achieved when the initial document set was retrieved using phrases, while the subjective adjective-based re-ranking was done using single terms (run UWopinion3). This suggests that subjective adjectives located close to any word from the query are useful indicators of the presence of opinions expressed about the query concept.

References

[1] Büttcher S. and Clarke C.L.A. Indexing Time vs. Query Time Trade-offs in Dynamic Information Retrieval Systems. In Proceedings of the 14th ACM Conference on Information and Knowledge Management, Bremen, Germany, November 2005.

[2] Hatzivassiloglou V. and McKeown K. R., Predicting the semantic orientation of adjectives. In Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, pp. 174-181, 1997.

[3] Spärck Jones K, Walker S. and Robertson S.E. A probabilistic model of information retrieval: development and comparative experiments. Parts 1 and 2. Information Processing and Management, 2000, 36(6), pp. 779-808, 809-840.