Improving Opinion Retrieval Based on Query-Specific Sentiment Lexicon

Conference Paper · April 2009

DOI: 10.1007/978-3-642-00958-7_76 · Source: DBLP

4 authors:

Seung-Hoon Na
Chonbuk National University

Yeha Lee
Pohang University of Science and Technology

Sang-Hyob Nam
Pohang University of Science and Technology

Jong-Hyeok Lee
Pohang University of Science and Technology

Improving Opinion Retrieval Based on Query-Specific Sentiment Lexicon

Seung-Hoon Na[1], Yeha Lee[2], Sang-Hyob Nam[2], and Jong-Hyeok Lee[2]

[1] National University of Singapore
nash@comp.nus.edu.sg
[2] POSTECH, South Korea
{sion,namsang,jhlee}@postech.ac.kr

Abstract. Lexicon-based approaches have been widely used for opinion retrieval due to their simplicity. However, no previous work has focused on the domain-dependency problem in opinion lexicon construction. This paper proposes simple feedback-style learning for query-specific opinion lexicon using the set of top-retrieved documents in response to a query. The proposed learning starts from the initial domain-independent general lexicon and creates a query-specific lexicon by re-updating the opinion probability of the initial lexicon based on top-retrieved documents. Experimental results on recent TREC test sets show that the query-specific lexicon provides a significant improvement over previous approaches, especially in BLOG-06 topics[1].

1 Introduction

Approaches for opinion retrieval should handle two types of scoring problems– 1) 'relevance scoring' of a document which is calculated based on topical relevance only, and 2) 'opinion scoring' of a document which is calculated based on how opinionatedly described a document is. Among these two problems, this paper focuses on 'opinion scoring' because it is not handled by previous retrieval models.

Previous approaches for opinion scoring are divided into two categories – 1) classification approach and 2) lexicon-based approach [2]. The lexical-based approach is preferred by most researchers, because of its simplicity and non-dependence on machine learning techniques.

However, a domain-independent lexicon such as SentiWordNet which the lexicon-based approaches have used widely is not sufficient to handle the opinion scoring problem, because there are many domain-specific opinionated words which cannot be derived from domain-independent words. Differently from domain-independent words, domain-specific ones can be created or derived from objects and attributes in corresponding domains. For example, a movie review would likely contain expressions of "interesting story", "truthfulness" and "high

[1] This work corresponds to one of the best results submitted by our team to TREC '08 Blog opinion track [1].

M. Boughanem et al. (Eds.): ECIR 2009, LNCS 5478, pp. 734–738, 2009.

performance of acting". A note-book review would likely contain "fast", "light" and "pretty", etc. Because of this domain-specific problem, the issue of whether or not a given document is opinionated should be resolved by using the domain-specific lexicon, in addition to domain-independent lexicon.

Unfortunately, it is known as a non-trivial problem to construct a separate opinion lexicon for each domain in opinion retrieval. First, the opinion retrieval belongs to the open-domain problem, thus all possible domains such as politics, entertainments, sports, science, people and food should be considered. Second, we should also obtain the corresponding domain-specific training data for every domain which is the collection of opinionated and non-opinionated documents or sentences, even though only a small number of available documents is available for a lot of domains.

To avoid these complicated issues, this paper proposes *feedback-style learning* for constructing a domain-specific lexicon, in which top-retrieved documents are regarded as the training data to learn opinion lexicon for *query-domain*, without prefixing the separate lexicon or constructing the training data for each domain. The feedback-style learning assumes that top-retrieved documents are a good resource for learning query-domain specific lexicon. We propose a simple and heuristic-based feedback-style learning. The proposed feedback-style learning first starts from the initial domain-independent lexicon, and then re-updates the initial model by observing top-retrieved documents according to how a given word frequently occurs in documents with high degrees of subjectivity, finally producing *query-specific opinion lexicon*.

2 Basic Opinion Retrieval Based on Lexicon-Based Approach

First, we start from briefly describing our basic lexicon-based approach. Without the loss of generality, we assume that our opinion retrieval system has a general opinion lexicon O, where O simply means the set of all possible opinion words. To generalize lexical-based approach, we consider probabilistic opinion lexicon where each opinion word w is assigned by a numeric value – $P(Subj|w)$ – the probability that word w belongs to opinionated status in a text. We simply call $P(Subj|w)$ *subjectivity* of w.

Now, suppose that query Q and document D are given. Let $tf(w;D)$, $len(D)$ and $ulen(D)$ be the term frequency of w in document D, the length of document D and the number of unique terms in a document D, respectively.

Generally, an opinion retrieval is summarized to the combination of two different types of scores – the *relevance score* of D ($relscore(Q,D)$) and the *opinion score* of D ($opinscore(Q,D)$) as follows:

$$score(Q,D) = (1-\alpha)\ relscore(Q,D) + \alpha\ opinscore(Q,D) \qquad (1)$$

The most simple metric for $opinscore(Q,D)$ is Avg, the average subjectivity of terms in document D as follows:

$$opinscore(Q,D) = \frac{\sum_w P(Subj|w) tf(w;D)}{len(D)} \qquad (2)$$

We extend the definition of term frequency to opinionated lexicon word, by introducing $tf(O;D)$ for the term frequency of opinion lexicon O in the document D. Since O is not a single term but a set, all term frequencies of opinionated words in O should be reflected to $tf(O;D)$. We regard that an occurrence of opinionated word w corresponds to the fuzzy count of $P(Subj|w)$. We use fuzzy count, since it will be more helpful to discriminate a strongly opinionated word from a weakly opinionated word. From this, we can define $tf(O;D)$ as follows:

$$tf(O;D) = \sum_{w \in O} P(Subj|w) tf(w;D) \tag{3}$$

Then, we connect the opinion scoring to retrieval modeling problem, by regarding O as additional pseudo single query term. Among several retrieval models, this paper considers the following BM25 model for $opinscore(Q,D)$ [3]:

$$opinscore_{BM25}(Q,D) = \frac{(k+1)tf(O;D)}{tf(O;D) + k\left(1 - b + b\frac{len(D)}{avglen}\right)} \tag{4}$$

where $avglen$ indicates the average length for all documents in a collection [3], and k and b are tuning parameters.

3 Feedback Approach for Learning Query-Specific Opinion Lexicon

Now we describe the proposed feedback-style learning of query-specific opinion lexicon. Initially, suppose that opinion lexicon is given by $P(Subj|w)$ obtained from domain-independent opinionated lexicon such as SentiWordNet, and that F is the set of top-retrieved documents in response to a given query. For each document D, we first estimate *document-level subjectivity* $P(Subj|D)$ from initial lexicon model, i.e. the probability of how much opinionated a given document D is, as follows:

$$subj(D) = \sum_{w \in D} P(Subj|w)/ulen(D)$$

$$P(Subj|D) = \frac{subj(D)}{max_{D \in F} subj(D)} \tag{5}$$

Next, the new probabilistic model of opinion lexicon – $P'(Subj|w)$ – is derived from these document-level subjectivities of documents as follows:

$$P'(Subj|w) = \sum_{D \in F} P(Subj|D) P(D|w) \tag{6}$$

where $P(D|w)$ indicates the probability that a document which contains word w is D. For simplicity, we assume that $P(D|w)$ is uniformly distributed on $F(w)$, the subset of feedback documents which contain w, as follows:

$$P(D|w) = \begin{cases} 1/|F(w)| & \text{if } w \in D \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

Note that $P(D|w)$ is not defined when word w does not appear on top-retrieved documents.

3.1 SentiWordNet as Initial Lexicon Model

This paper used SentiWordNet to define the initial lexicon model [4], and encoded $P(Subj|w)$ from the knowledge of SentiWordNet. SentiWordNet defines the degree of polarity for each sense s (WordNet synset) – $P(Neg|s)$ and $P(Pos|s)$ for negative and positive polarity, respectively. We define (approximately) $P(Subj|w)$ by $max_{s \in sense(w)} max(P(Neg|s), P(Pos|s))$ where $sense(w)$ is the set of all WordNet synsets corresponding to w.

3.2 Using Passage Context

Generally, a document consists of several topics, rather than presenting a single topic, so that not all parts but some parts of a relevant document likely to be actually relevant to query. We consider the passage-context to more accurately learn the query-domain opinion lexicon, instead of using the whole-document context. Our approach to extract passage-level context consists of two steps. In the first passage retrieval step, we adopted the completely-arbitrary passage retrieval [5], finding the *most relevant passage* of each document to the given query. In the second passage extension step, the most relevant passage is further extended by enlarging its context by maximally L length for forward and backward directions, respectively [6]. This extended best passage is used as the context for calculating $P(Subj|D)$ and $P(D|w)$ for Eq. (6).

4 Experimentation and Conclusion

We performed retrieval runs on standard TREC blog test collection consisting of 3,215,171 permalinks [2]. The evaluation was performed separately on two query sets of BLOG-06 (Q850∼Q900) and BLOG-07 (Q901∼Q950). We used two different fields for a query topic – T (using only title field) and TD (using title and description fields). We considered MAP (Mean Average Precision) as evaluation measure.

We first generated top 3000 documents by using the baseline retrieval method of section 2. The opinion retrieval module re-ranked them to generate top 1000 documents, according to opinion scoring module. The following lists up the performed methods for opinion retrieval module:

1) **Baseline:** The baseline retrieval method without any opinion scoring.
2) **Doc-Avg:** The average subjectivity for $opinscore(Q, D)$ (Eq. (2)).
3) **Doc-Okapi-len:** The okapi retrieval model for $opinscore(Q, D)$ (Eq. (4)). The best parameters for k and b were selected and fixed to all queries.
4) **Psg-Okapi-len:** same as Doc-Okapi-len, except for using the extended best passage of each document for calculating $opinscore(Q, D)$.
5) **Psg-QS-Okapi-len:** Same as Psg-Okapi-len, except for using query-specific lexicon (section 3.2).

Table 1. Performances of opinion retrieval of five methods. Bold face numbers indicate the best run in each topic and test set.

Method	BLOG-06		BLOG-07	
	T	TD	T	TD
Baseline	0.2647	0.3022	0.3757	0.3784
Doc-Avg	0.2976	0.3263	0.4172	0.4125
Doc-Okapi-len	0.3041	0.3296	0.4393^{α}	0.4240^{α}
Psg-Okapi-len	0.3073^{α}	0.3313	0.4296^{α}	0.4207^{α}
Psg-QS-Okapi-len	$\mathbf{0.3159}^{\alpha\beta}$	$\mathbf{0.3471}^{\alpha\beta}$	$\mathbf{0.4399}^{\alpha\beta}$	$\mathbf{0.4248}^{\alpha\beta}$

Table 1 shows the performances of five opinion retrieval methods. To check whether or not a given method shows statistically significant improvement over Doc-Avg, we applied Wilcoxon signed-rank test and marked $^{\alpha}$ beside the performance number. Additionally, we marked $^{\beta}$ when Psg-OS-Okapi-len or Psg-Okapi-len significantly improves Doc-Okapi-len. All statistical tests are evaluated at 99% confidence-level. Note that Psg-QS-Okapi-len is only the method which shows significant improvements over Doc-Avg on every topics and test set. Psg-QS-Okapi-len makes significant improvement over Doc-Okapi-len in BLOG-06, but only slight improvement in BLOG-07. When a fully-probabilistic approach is deployed for query-specific learning, we believe that an significant improvement can be also obtained in BLOG-07.

Acknowledgement. This work was supported in part by MKE & IITA through IT Leading R&D Support Project and also in part by the BK 21 Project in 2008. We thank the reviewers for their helpful comments.

References

1. Lee, Y., Na, S.H., Kim, J., Nam, S.H., Lee, J.H.: KLE at TREC 2008 Blog Track: Blog Post and Feed Retrieval. In: TREC 2008 (2008)
2. Ounis, I., de Rijke, M., Macdonald, C., Mishne, G., Soboroff, I.: Overview of the TREC-2006 Blog Track. In: TREC 2006 (2006)
3. Robertson, S.E., Walker, S., Beaulieu, M.: Okapi at TREC-7: automatic ad hoc, filtering, vlc and interactive. In: TREC-7, pp. 253–264 (1999)
4. Esuli, A., Sebastiani, F.: SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In: LREC 2006 (2006)
5. Na, S.H., Kang, I.S., Lee, Y.H., Lee, J.H.: Completely-Arbitrary Passage Retrieval in Language Modeling Approach. In: Li, H., Liu, T., Ma, W.-Y., Sakai, T., Wong, K.-F., Zhou, G. (eds.) AIRS 2008. LNCS, vol. 4993, pp. 22–33. Springer, Heidelberg (2008)
6. Na, S.H., Kang, I.S., Lee, Y.H., Lee, J.H.: Applying Complete-Arbitrary Passage for Pseudo-Relevance Feedback in Language Modeling Approach. In: Li, H., Liu, T., Ma, W.-Y., Sakai, T., Wong, K.-F., Zhou, G. (eds.) AIRS 2008. LNCS, vol. 4993, pp. 626–631. Springer, Heidelberg (2008)