

Improve the Effectiveness of the Opinion Retrieval and Opinion Polarity Classification

Wei Zhang

Live Search
Microsoft Corporation
Redmond, WA 98052, USA
wzhan@microsoft.com

Lifeng Jia, Clement Yu

Department of Computer Science
University of Illinois at Chicago
Chicago, IL 60607, USA
{ljia,yu}@cs.uic.edu

Weiyi Meng

Department of Computer Science
Binghamton University
Binghamton, NY 13902, USA
meng@cs.binghamton.edu

ABSTRACT

Opinion retrieval is a document retrieving and ranking process. A relevant document must be relevant to the query and contain opinions toward the query. Opinion polarity classification is an extension of opinion retrieval. It classifies the retrieved document as positive, negative or mixed, according to the overall polarity of the query relevant opinions in the document. This paper (1) proposes several new techniques that help improve the effectiveness of an existing opinion retrieval system; (2) presents a novel two-stage model to solve the opinion polarity classification problem. In this model, every query relevant opinionated sentence in a document retrieved by our opinion retrieval system is classified as positive or negative respectively by a SVM classifier. Then a second classifier determines the overall opinion polarity of the document. Experimental results show that both the opinion retrieval system with the proposed opinion retrieval techniques and the polarity classification model outperformed the best reported systems respectively.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *retrieval models, selection process*. I.2.7 [Artificial Intelligence]: Natural Language Processing – *text analysis*.

General Terms

Algorithms, Performance, Experimentation.

Keywords

Opinion Retrieval, Blog Retrieval, Information Retrieval.

1. INTRODUCTION

In opinion retrieval, a relevant document must be relevant to the query and contain opinions about the query. Significant improvement of the opinion retrieval effectiveness has been reported [8]. As a rather new research area, analyzing and incorporating more methods into the existing model to achieve higher retrieval effectiveness is of high interest. After the opinion retrieval, it is natural to consider the positive and the negative polarities of the query related opinions in the retrieved documents. The “polarity classification” task of TREC 2007 blog track [4] defines that a document has a positive (negative) label if the query

related opinions in this document are positive (negative). The document receives a mixed polarity label if it contains both positive and negative relevant opinions. Assigning the opinion polarity label at the document level is as important as doing it at the sentence level, as it gives a user an overview of the opinion orientation of a document. The user does not need to go to the fine detail if not interested to the opinion orientation.

This paper focuses on two objectives. First, it improves the opinion retrieval effectiveness of a strong baseline system [8] by studying the methods such as removing part-of-speeches restriction, using more opinionated features, accurately determine query-related opinions, adding spam filter, and utilizing the abbreviations of query concepts. The second objective is to present a two-stage model to solve the opinion polarity classification problem. The proposed model contains two components. (1) A sentence-level opinion classifier classifies a query related opinionated sentence as either positive or negative. (2) A document-level opinion classifier takes the output of the first classifier as input, and assigns a positive, negative or mixed opinion polarity label to the whole document.

2. RELATED WORKS

Zhang et al. proposed an opinion retrieval model that combines factual retrieval and opinion identification [8]. Mei et al. proposed a topic sentiment mixture model to model the facts and the opinions in blogs [9].

To identify and classify the subjective, Esuli and Sebastiani [1] classified the orientation of a term based on its dictionary glosses. Whitelaw et al. [6] used “appraisal groups” features for sentiment classification.

In the document-level opinion classification, Pang and Lee [5] classified movie reviews into multiple categories. Ku et al. [2] summarized blogs by their major topics and the related opinions.

3. IMPROVE OPINION RETRIEVAL EFFECTIVENESS

The opinion retrieval model in [8] is used as the baseline. This model has (1) A topic retrieval component. (2) An opinion detection component. (3) A ranking component. The details can be found in [8]. In this paper, we propose the following methods to improve the opinion retrieval effectiveness of the baseline.

- (1) Do not apply part-of-speech restriction in feature selection
- (2) Obtain more subjective features from the positive and negative training data by using chi-square test.

Copyright is held by the author/owner(s).
CIKM'08, October 26–30, 2008, Napa Valley, California, USA.
ACM 978-1-59593-991-3/08/10.

- (3) Refine the NEAR operator by allowing phrase searching in the text window.
- (4) Add spam filters to remove documents with long sentences or pornographic contents.
- (5) Add an “in-document-abbreviation-recognition” method to find more abbreviations of concepts for the NEAR operator.

4. TWO-STAGE POLARITY CLASSIFICATION

The sentence-level opinion polarity classifier is similar to the subjective/objective sentence classifier of [8]. But now the SVM classifier is trained by the positive and negative features obtained from positive and negative review training data. Given a retrieved document from the opinion retrieval system, this classifier assigns each query relevant opinionated sentence in this document a positive/negative label and a score.

Based on the sentence-level opinion orientation information, the document-level opinion classifier labels the whole document as positive (or negative), if it only contains positive (or negative) query relevant opinions, or mixed if both sufficient positive and negative opinions are found. This classifier utilizes a set of sentence-level features, and is implemented as a decision tree classifier. Given a document d , the features are

- (1) The number of the positive relevant sentences in d .
- (2) The number of the negative relevant sentences in d .
- (3) The sum of the classification scores of the sentences in d that are classified to be positive relevant.
- (4) The sum of the classification scores of the sentences in d that are classified to be negative relevant.
- (5) Average score of classified positive relevant sentences in d .
- (6) Average score of classified negative relevant sentences in d .
- (7) The ratio of the number of the classified positive relevant sentences in d , to the number of the classified negative relevant sentences in d .
- (8) The ratio of the sum of the scores of the classified positive relevant sentences in d to the sum of the scores of the classified negative relevant sentences in d .

5. EXPERIMENTAL RESULTS

5.1 Improve Opinion Retrieval Effectiveness

This experiment takes the opinion retrieval system in [8] as the baseline. The five proposed techniques are incorporated into the baseline one by one. This results in six system configurations. Each configuration is tested by TREC 2006 [3] and 2007 [4] title-only blog queries respectively on the TREC Blog data set [3]. The results are shown in Table 1 in terms of MAP score.

5.2 Opinion Polarity Classification

This experiment takes the system of [7] as the baseline, as it reported the highest classification R-accuracy score on TREC 2007 blog queries [4]. To test our model, the TREC 2006 data is used to train the classifier in order to test the TREC 2007 data, and vice versa. The baseline system does not require classifier training. The R-accuracy scores of the baseline and the proposed model on 2006 and 2007 blog queries are shown in Table 2.

Table 1. MAP scores of different system configurations on TREC 2006 and 2007 query set

	Configurations	2006 query	2007 query
(1)	Baseline	0.2837	0.4086
(2)	(1) + No POS restriction	0.2880	0.4130
(3)	(2) + More subjective features	0.2888	0.4164
(4)	(3) + Refined NEAR operation	0.2925	0.4280
(5)	(4) + Removing spam	0.2987	0.4344
(6)	(5) + Using abbreviations	0.3009	0.4425
	(6) vs. (1), MAP increases	6.06%	8.30%

Table 2. The R-Accuracy scores of the proposed model and the baseline model on TREC 2006 and 2007 data sets

Data Set	Baseline	Proposed Model	Increase
TREC 06	0.1595	0.1619	1.50%
TREC 07	0.2322	0.2517	8.40%

6. CONCLUSIONS

This paper presents several techniques that improve the retrieval effectiveness of an opinion retrieval system. It also proposes a two-stage opinion polarity classification model. Both the improved opinion retrieval system and the polarity classification model outperformed the state-of-art systems respectively on the same data sets. In the future, we plan to study the spam detection techniques to further improve the retrieval effectiveness.

7. ACKNOWLEDGMENTS

This research is supported in part by NSF grant IIS-0738652 and NSF grant IIS-0738727.

8. REFERENCES

- [1] A Esuli and F Sebastiani. Determining the semantic orientation of terms through gloss analysis. In *Proc. of CIKM*. 2005.
- [2] L-W Ku, L-Y Lee, T-H Wu, and H-H Chen. Major topic detection and its application to opinion summarization. In *Proc. of SIGIR*. 2005
- [3] I Ounis, M de Rijke, C Macdonald, G Mishne, and I Soboroff. Overview of the TREC-2006 Blog Track. In *TREC* 2006.
- [4] I Ounis, C Macdonald and I Soboroff. Overview of the TREC-2007 Blog Track. In *TREC*. 2007.
- [5] Bo Pang and Lillian Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proceedings of ACL*. 2005.
- [6] C Whitelaw, N Garg, and S Argamon. Using appraisal groups for sentiment analysis. In *Proc. of CIKM*. 2005
- [7] Wei Zhang and Clement Yu. UIC at TREC 2007 Blog Track. In *TREC*. 2007.
- [8] Wei Zhang, Clement Yu and Weiyi Meng. “Opinion Retrieval from Blogs”. In proceedings of CIKM. 2007.
- [9] Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su and Chengxiang Zhai. Topic sentiment mixture: modeling facets and opinions in weblogs. In *proc. of WWW*. 2007.