12-2011

An effective approach for topicspecific opinion summarization

Binyang LI

Lanjun ZHOU

Wei GAO
Singapore Management University, weigao@smu.edu.sg

Kam-Fai WONG

Zhongyu WEI

Citation

LI, Binyang; ZHOU, Lanjun; GAO, Wei; WONG, Kam-Fai; and WEI, Zhongyu. An effective approach for topicspecific opinion summarization. (2011). *Proceedings of the Seventh Asian Information Retrieval Societies Conference*. 398-409. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/4591

An Effective Approach for Topic-Specific Opinion Summarization

Binyang Li, Lanjun Zhou, Wei Gao, Kam-Fai Wong, and Zhongyu Wei

Department of Systems Engineering and Engineering Management
The Chinese University of Hong Kong Shatin, NT, Hong Kong, China
Key Laboratory of High Confidence Software Technologies Ministry of Education, China
{byli,ljzhou,wgao,zywei,kfwong}@se.cuhk.edu.hk

Abstract. Topic-specific opinion summarization (TOS) plays an important role in helping users digest online opinions, which targets to extract a summary of opinion expressions specified by a query, i.e. topic-specific opinionated information (TOI). A fundamental problem in TOS is how to effectively represent the TOI of an opinion so that salient opinions can be summarized to meet user's preference. Existing approaches for TOS are either limited by the mismatch between topic-specific information and its corresponding opinionated information or lack of ability to measure opinionated information associated with different topics, which in turn affect the performance seriously. In this paper, we represent TOI by word pair and propose a weighting scheme to measure word pair. Then, we integrate word pair into a random walk model for opinionated sentence ranking and adopt MMR method for summarization. Experimental results showed that salient opinion expressions were effectively weighted and significant improvement achieved for TOS.

Keywords: Topic-specific opinion summarization, topic-specific opinionated information, word pair, MMR.

1 Introduction

With the development of Web 2.0, people have become interested in expressing their personal opinions through online tools. There are a great amount of opinions widely spread from the comments on the health reform to the evaluations on a feature of a consumer product. In practice, people would like to focus on the summary of opinion expressions with their own preference to make decision [1]. For instance, users would like to give a query, "what are the opinions on X?" to express their preference of X, where X could not only be a feature of a product, but also be a target of a general domain opinion. Therefore, it is significant to study topic-specific opinion summarization(TOS) to meet the user's personal preference.

In Example 1, there are three opinion expressions tagged in bold in sentence (a). If a query is given about a *computer game*, traditional summarization regards only topical relevance, such as "*game*", "*operation*", "*screen*" to be the information need. For TOS, however, it is supposed to take both topic-specific information and

M.V.M. Salem et al. (Eds.): AIRS 2011, LNCS 7097, pp. 398–409, 2011.

opinionated information into consideration, e.g., "*game is very small*". In this paper, we define user's information need of TOS as topic-specific opinionated information (TOI), i.e. the opinion expressions about the user's query.

> Example 1:
> (a) *The* [***game is very small***], [***operation is very flexible***], *and* [***screen is beautifully smooth***].
> (b) *The* [***screen is small but adequate***].

One of the fundamental problems in TOS is how to effectively represent the user's information need so as to evaluate and summarize salient opinion expressions.

Previous methods using KL-divergence [2] or feedback-style learning [3] have the limitation that TOI is represented by one single word. In practice, one single word can hardly represent both topic-specific information and opinionated information at the same time, especially for those domain-independent sentiment words that barely represent topic-specific information, e.g., "*small*" in Example 1. [4, 5] proposed to express topic-specific information and opinionated information by topic-specific words and sentiment words, respectively. However, they regarded the document as *bag-of-word*, and neglected the contextual information, which means word-based representation cannot hold the associative information between topic-specific information and opinionated information in individual opinion expression and lead to a mismatch. In Example 1, although topic-specific information and opinionated information of three opinion expressions can be represented separately, the associative information between them is lost due to lack of contextual information. In an extreme situation, the fake opinion expression "*small screen*" will be selected as the salient opinion expression.

Li et al. proposed to adopt word pair to represent TOI [6]. A word pair is constructed by a sentiment word together with its corresponding topic-specific word. The sentiment word represents opinionated information, i.e. the opinion, and the corresponding topic-specific word represents topic-specific information, i.e. target (also refer to the topic-specific word in this paper). With the help of pairwise representation, the contextual information between the opinion and its corresponding target could be maintained. Nonetheless, Li et al. neglected to measure the variant associations between the sentiment word and the topic-specific word within different word pair. In [6], a unified trade-off parameter was introduced to balance the topic-specific information and opinionated information for all word pairs. In practice, it is inadequate to describe the distinct association of individual word pair, because one sentiment word is supposed to modify different targets in different opinion expressions. In Example 1, both sentence (a) and sentence (b) include the sentiment word "*small*" which associated with different topic-specific words ("*game*" in (a) and "*screen*" in (b)). Intuitively, "*small*" is a domain-independent sentiment word, and it should be dynamically assigned different weights when modifying different targets. Therefore, we argue that the trade-off parameter should be estimated for each word pair to describe the distinct association.

In this paper, we also follow the TOI representation by word pair. According to the above analysis of word pair, we first propose an effective weighting scheme from the perspective of information gain by selecting the word pair to measure both topic-specific words and sentiment words, and then provide an individual associative score for each word pair. Thus, the TOI of individual opinion expression is able to be measured. Finally, we integrate word pair into a random walk model for sentence ranking and adopt maximal marginal relevance (MMR) method to generate the topic-specific opinion summary.

To investigate the effectiveness of our approach, experiments were made based on the TAC2008 and OpQA benchmark datasets. Significant improvements over the best run in TAC 2008 and those models with word-based representation were shown in this paper.

In the remainder of this paper, we first present pairwise representation of topic-specific opinionated information together with a new weighting scheme in Section 2. We then integrate word pair into a random walk model for sentence ranking and generate opinion summary by using MMR method in Section 3. In Section 4, we will show the experimental results. We review the related work in Section 5. Finally, this paper is concluded and the future work is suggested in Section 6.

2 Representation of Topic-Specific Opinionated Information

Topic-specific opinion summarization (TOS) was first proposed in the Text Analysis Conference (TAC) 2008, and the objective is to extract an informative summary of opinion expressions about a given query, as found in a document collection [7]. Different from traditional topic-specific summarization that concentrates only on the topic-specific information, TOS concerns on the topic-specific opinionated information (TOI). More precisely, the TOI of an opinion expression is supposed to contain the following attributes: opinion (i.e. opinionated information), holder, target (i.e. topic-specific information) and polarity [8]. In this section, we will first describe how to express TOI by pairwise representation, word pair. Then a new weighting scheme is introduced for measuring the pairwise representation.

2.1 Pairwise Representation

Without loss of generality, we assume that there is a document set $\mathcal{D}$ ($\mathcal{D} = \{d_1, d_2, d_3, \cdots, d_n\}$) that includes a set of sentences $\mathcal{S} = \{s_1, s_2, s_3, \cdots, s_N\}$, and a user generated query $\mathcal{Q} = \{q_1, q_2, q_3, \cdots, q_z\}$, where $q_1, q_2, q_3, \cdots, q_z$ are the key words. TOS aims at extracting an informative summary of $\mathcal{S}'$ ($\mathcal{S}' \subseteq \mathcal{S}$) with opinion expressions from $\mathcal{D}$ about the $\mathcal{Q}$.

In order to represent TOI, we need to consider all the attributes of an opinion expression together with the associations between these attributes. We also utilize topic-sentiment word pair [6] in this paper. The notion of word pair was first proposed for opinion retrieval to capture the contextual information between the opinion and its corresponding target [6]. Since most opinion holders are implicit to be the author in the blogosphere, we do not take opinion holder attribute into consideration.

Definition 1: topic-sentiment word pair p_{ij} consists of two elements, one represents the opinion, and the other one represents the modified target, $p_{ij} = \{< t_i, o_j > | t_i \in V_t,\ o_j \in V_o\}$.

V_t is the topic-specific words collection with all candidate targets, and the target reflects the preference of user by the query matching. V_o is the sentiment word lexicon which is used to express opinions. We maintain the semantic information between the topic-specific word and the sentiment word by pairwise representing. Thus, TOI of an opinion expression is represented by a word pair. We assume that for one query the candidate targets and opinions are in V_t, and V_o, respectively. The total number of the word pair is $|p_{ij}| = m \times M$, $(|V_t| = m\ , |V_o| = M)$.

In practice, the weight of a sentiment word may differ from variant targets. In the following section, we will describe a weighting scheme to measure word pair from the perspective of information gain, and introduce a method by normalizing Point-wise Mutual Information (PMI) between the sentiment word and the corresponding topic-specific word to compute the associative score for individual opinion expression.

2.2 Weighting Scheme for Word Pair

Previous weighting schema would like to capture both topic-specific information and opinionated information by one single word. It either assigns a *relevance* weight to the sentiment word, such as using the distribution divergence from the query words [9], or integrates the sentiment weight into query word, e.g., computes the Point-wise Mutual Information (PMI) score between sentiment word and the target combined with *tf-idf* value of query words [10]. In this work, we regard the topic-specific information and opinionated information of an opinion expression to be represented by topic-specific word and sentiment word of a word pair, respectively. Therefore, we measure both topic-specific words and sentiment words by computing the information gain in selecting the word pair.

In TOS, both sentiment words and topic-specific words are considered as informative content words, and described as "*term*" (denoted by w). Additionally, we would like to concentrate on the granularity of sentence rather than document.

We can compute term weight of w, $TW(w)$ by Equation (1):

$$TW(w) = -\log_2 P(w)/P'(w) \tag{1}$$

For simplicity, we assume that any term w_i follows *Poisson* distribution. (discuss other distributions in Section 4) $P(w)$ on the whole set of words with the parameter $\lambda_w = |w|/|S|$ ($|S|$ is the total number of words in S), while it also follows another *Poisson* distribution $P_1{}'(w)$ on the set of sentences including w with the parameter $\lambda'_w = |w|/|S(w)|$ ($|S(w)|$ is the total number of words in the sentences including w). Obviously $\lambda_w \le \lambda'_w$.

We measure each term according to its distributions between the sentence set it occurs and the whole sentence set. In other words, we weigh sentiment word by computing the gain in selecting a sentence containing the sentiment word.

Recall Example 1, one sentiment word (resp. target) may be assigned different weights when associated with different targets (resp. sentiment words). Therefore, there is a must to embody the different associations between sentiment words and topic-specific words.

Previous works [4, 5, 6] focus on using a unified parameter to express variant combinations between topic-specific information and opinionated information. It is inadequate to express the variant of associations even to one specific domain. We propose to compute an associative score (also referred as trade-off parameter in this paper) for each individual association between the sentiment word and the topic-specific word.

Inspired by the fact that Mutual Information is a measurement to assess how two words are associated, and achieves better performance in [12], we therefore utilize mutual information to estimate the trade-off parameter for both sentiment words and the target words.

In our method, for each word t_i in V_t, we compute its mutual information scores for all words in V_o and normalize the scores. Informally, mutual information compares the probability of observing t_i and o_j together (the joint probability) with the probabilities of observing t_i and o_j independently. The mutual information between words t_i and o_j are calculated as follows:

$$I\left(X_{t_i}; X_{o_j}\right) = \sum_{X_{t_i}=0,1} \sum_{X_{o_j}=0,1} \log \frac{p(X_{t_i}, X_{o_j})}{p(X_{t_i})p(X_{o_j})} \quad (2)$$

where X_{t_i} and X_{o_j} are binary variables indicating whether t_i or o_j is present or absent.

The parameters are estimated as follows:

$$p(X_{t_i} = 1) = c(X_{t_i} = 1)/N$$
$$p(X_{t_i} = 0) = 1 - p(X_{t_i} = 1)$$
$$p\left(X_{o_j} = 1\right) = c\left(X_{o_j} = 1\right)/N$$
$$p\left(X_{o_j} = 0\right) = 1 - p\left(X_{o_j} = 1\right)$$
$$p\left(X_{t_i} = 1, X_{o_j} = 1\right) = \frac{c\left(X_{t_i} = 1, X_{o_j} = 1\right)}{N}$$
$$p\left(X_{t_i} = 1, X_{o_j} = 0\right) = \frac{c(X_{t_i} = 1) - c\left(X_{t_i} = 1, X_{o_j} = 1\right)}{N}$$
$$p\left(X_{t_i} = 0, X_{o_j} = 1\right) = \frac{c\left(X_{o_j} = 1\right) - c\left(X_{t_i} = 1, X_{o_j} = 1\right)}{N}$$
$$p\left(X_{t_i} = 0, X_{o_j} = 0\right) = 1 - p\left(X_{t_i} = 1, X_{o_j} = 1\right) - p\left(X_{t_i} = 0, X_{o_j} = 1\right) - p\left(X_{t_i} = 1, X_{o_j} = 0\right)$$

where $c(X_{t_i} = 1)$ and $c\left(X_{o_j} = 1\right)$ are the numbers of sentences containing word t_i and o_j, respectively, $c\left(X_{t_i} = 1,\ X_{o_j} = 1\right)$ is the number of sentences that contain both t_i and o_j, and N is the total number of sentences in the collection. We then normalize the mutual information score to obtain a trade-off parameter $\mu_{mi}(o_j|t_i)$ to balance the weight of t_i when associated with o_j:

$$\mu_{mi}(o_j|t_i) = \frac{I(X_{t_i}; X_{o_j})}{\sum_{o \in \mathrm{Vo}} I(X_{t_i}; X_o)}$$

$$\mu_{mi}(t_i|o_j) = \frac{I(X_{t_i}; X_{o_j})}{\sum_{t \in \mathrm{V_t}} I(X_{o_j}; X_t)}$$

$\mu_{mi}(t_i|o_j)$ is computed in the same way to balance the weight of o_j when associated with t_i. The probability would be higher if the two words co-occur with each other more frequently.

After estimating the associative score between the two elements of a word pair, we can assign the weight to individual word pair p_{ij}.

$$w_{p_{ij}} = \mu_{mi}(t_i|o_j) \times TW(t_i) + \mu_{mi}(o_j|t_i) \times TW(o_j) \tag{3}$$

where $TW(t_i)$ and $TW(o_j)$ are the term weights of t_i and o_j, which can be computed from Equation (1).

As to those word pairs with a negation operator around, an alternative would be to rewrite the individual word pair p_{ij} as $\neg p_{ij}$. Since the negation operator only shifts the polarity of the word pair, we assign $\neg p_{ij}$ with the same weight as p_{ij}.

3 Word Pair Based TOS

In this section, we first integrate word pair into a random walk model [14], for sentence ranking. Then, we utilize MMR method to generate a summary.

3.1 PageRank Based on Word Pair

In Section 2, we introduce a weighting scheme for measuring individual word pair. According to the definition of word pair, it is intuitively that the sentence with word pair representing salient opinion expression should be assigned a relatively high weight. We, therefore, consider the global information of word pair for sentence ranking by the recursive procedure in the random walk model, PageRank.

One of our objectives is to investigate the effectiveness of our proposed weighting scheme for word pair, so we compute the similarity between sentences according to the weighted word pair. Moreover, in our approach, we do not explicit divide sentiment words into domain-dependent and domain-independent, but use the corresponding target as an indicator. This will weaken the effect of domain-independent sentiment words. In order to correct the opinionated information of a domain-independent sentiment of a word pair, we utilize synonym dictionary SentiWordNet [13]. We choose the sentiment word with the highest PMI score over a topic-specific word in Equation (2) as the cue word and consider all synonyms of the cue word together with the corresponding target to be the same word pair.

We define a PageRank model that has sentences to be summarized as nodes and edges placed between two sentences that are similar to each other.

We can then score all the sentences based on the expected probability of a random walker visiting each sentence. We use the short-hand $P(s_u|s_v)$ to denote the probability of being at node s_u at a time t given that the walker was at s_v at time $t-1$. The jumping probability from node s_v to node s_u is given by:

$$P(s_u|s_v) = \frac{sim(s_v,s_u)}{\sum_{s_{u'} \in S \backslash s_v} sim(s_v,s_{u'})} \tag{4}$$

where sim is a similarity function defined on two sentence/excerpt nodes based on the word pair they contain.

$$sim(s_u, s_v) = \frac{\sum_{p_{ij} \in s_u, s_v} w_{p_{ij}}^{s_u} \cdot w_{p_{ij}}^{s_v}}{\sqrt{\sum_{p_{ij} \in s_u} (w_{p_{ij}}^{s_u})^2} \times \sqrt{\sum_{p_{ij} \in s_v} (w_{p_{ij}}^{s_v})^2}} \quad (5)$$

The saliency score $Score(s_u)$ for sentence s_u can be calculated by mixing query similar score and scores of all other sentences linked with it as follows:

$$Score(s_u) = \gamma \sum_{v \neq u} Score(s_v) \cdot P(s_u|s_v) + (1-\gamma) sim'(s_u|Q) \quad (6)$$

where $sim'(s_u|Q) = \frac{sim(s_u|Q)}{\sum_{k=1}^{N} sim(s_k|Q)}$

$$sim(s_u|Q) = \frac{\sum_{w_{t_i} \in s_u, Q} w_{t_i}^{s_u} \cdot w_{t_i}^{Q}}{\sqrt{\sum_{w_{t_i} \in s_u} (w_{t_i}^{s_u})^2} \times \sqrt{\sum_{w_i \in Q} (w_{t_i}^{Q})^2}} \quad (7)$$

where $w_{t_i}^{s_u}$ $w_{t_i}^{Q}$ are the weights of t_i in s_u, and Q, respectively.

Finally, all the sentences will rank according to the saliency score. As for each query, we choose a number of sentences with weights higher than a threshold as candidate set R for TOS.

3.2 Summary Generation

In order to generate a summary, we adopt maximal marginal relevance (MMR) method and incrementally add the top ranked sentences from R into the answer set.

$$MMR = Arg \max_{s_u \in R \setminus S'} \left[\theta\big(sim(s_u|Q)\big) - (1-\theta) \max_{s_v \in S'} sim(s_u, s_v) \right]$$

R is the ranked list of sentences retrieved by the PageRank model in Section 3.1, given the document set D and a query Q. We set the relevance threshold δ, below which it will not be regarded as candidate sentences. S' is the subset of sentences in R already selected; $R \setminus S'$ is the set difference, i.e., the set of as yet unselected sentences in R. We compute $sim(s_u, s_v)$ and $sim(s_u|Q)$ and by Equation (5) and Equation (7), respectively in Section 3.1. The parameter θ lying in $[0,1]$ controls the relative importance given to "relevance" versus redundancy. As different users with different information needs may require a totally different summary, especially for TOS, one of the attractive points of MMR is by setting the value of parameter θ, it can particularly generate summaries according to a user's need. In the experiment, we set $\theta = 0.5$ to balance the novelty and the relevance.

4 Evaluation

4.1 Experiment Setting

4.1.1 Benchmark Datasets

Our experiments are based on two benchmark datasets for topic-specific opinion summarization, TAC2008 and OpQA.

TAC2008 dataset is the benchmark data set for the topic-specific opinion summarization track in the Text Analysis Conference 2008 (TAC2008), which contains 87 squishy opinion questions. The initial topic words for each question are also provided. Summarizations for all queries must be retrieved from the TREC Blog06 collection [15], which consists of review and blog data. The top 50 documents were retrieved for each query.

The Opinion Question Answering (OpQA) corpus consists of 98 documents appeared in the world press between June 2001 and May 2002. The documents covered four general topics, and 30 questions were given. [16]

4.1.2 Sentimental Lexicon and Topic Collection

In our experiment, we use SentiWordNet as the sentiment lexicon. SentiWordNet is a popular lexical resource for opinion mining, which consists of 4800 negative sentiment words and 2290 positive sentiment words. For each sentiment word, SentiWordNet also provides its synonyms.

In order to acquire the collection of topic terms, we adopt two expansion methods, dictionary- based method and pseudo relevance feedback method [6].

4.1.3 TOS Approaches for Comparison

To demonstrate the effectiveness of pairwise representation for TOS, we compared it with the following models:

(1) Baseline 1: This model [18] achieved the best run in TAC2008 opinion summarization task. We treated it as Baseline 1 in the experiment.

(2) OPM-1: This model was proposed for opinion question and answering, which achieved 2% improvement over the best run in TAC2008 Opinion QA track [19].

(3) OPM-2: This model was similar with OPM-1, but use PageRank model for sentence ranking instead.

(4) GOSM: This model was originally designed for opinion retrieval, and it adopted pairwise representation of TOI. GOSM adopted "*relevance*" measurement for sentiment word and utilized a uniform parameter to balance topic-specific information and opinionated information. We re-designed GOSM to deal with TOS by using Pair-based HITS model so that we could compare the effectiveness of different weighting schema for word pair [6].

(5) PPM: our proposed approaches.

Additionally, in our experiments, we will also investigate the performance of sentence retrieval with different probability models. We used the metrics in the Text Retrieval Conference (TREC), which are average precision (AvPr), R-precision (R-Pre) and precision at 10 sentences (P@10).

4.2 Performance Evaluation

4.2.1 Parameter Tuning

In our proposed approach, there are two parameters θ and γ. θ is a user-defined parameter according to the specific need. In our experiment, we set the parameter $\theta = 0.5$ to balance the novelty and the relevance.

We studied how the parameter γ (in Equation (6)) influenced the performance of sentence ranking in both TAC2008 and OpQA datasets. The results are given in Fig. 1.

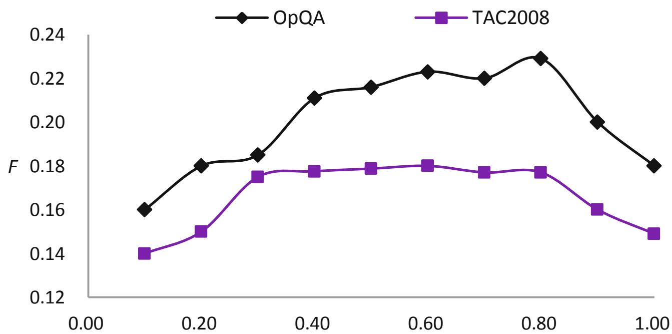

Fig. 1. Pair-based PageRank Performance with varying parameter γ on TAC2008 and OpQA

Best F value was achieved, when γ was set around 0.8 in both TAC2008 and OpQA datasets. Therefore, in the following experiments, we set $\gamma = 0.8$.

4.2.2 Comparisons on Sentence Ranking

In our evaluation, we also tested the performance of sentence ranking of other probability models, including *tf-idf* model and Bose-Einstein model. We used the metrics in the Text Retrieval Conference 10 (TREC), which are average precision (AvPr), R-precision (R-Prec) and precision at 10 sentences (P@10). In our experiment, we created the judgment through pooling method.

The experimental results based on these metrics are shown in Table 1.

Table 1. Comparison of sentence ranking on OpQA and TAC2008 datasets

Dataset	Probability	Metrics		
		AvPr	R-Prec	P@10
OpQA	Poisson	0.212	0.233	0.408
	tf-idf	**0.229**	0.230	0.397
	Bose-Einstein	0.208	**0.245**	**0.421**
TAC 2008	Poisson	**0.180**	0.206	0.361
	tf-idf	0.177	0.198	0.354
	Bose-Einstein	0.175	**0.212**	**0.369**

Table 1 showed that Bose-Einstein model achieved best R-Prec and P@10 on OpQA and TAC2008 datasets. Thus, we chose Bose-Einstein model for further evaluation.

4.2.3 Comparisons on TOS

We were also interested in the performance comparison with the other models for TOS.

Table 2. Comparison of TOS on OpQA and TAC2008 datasets

Data set	Approaches	Measurements		
		Precision	Recall	F(3)
OpQA	Baseline 1	0.280	0.356	0.325
	OPM-1	0.274	0.368	0.343
	OPM-2	0.281	0.354	0.362
	GOSM	0.286	0.360	0.379
	PPM	0.276	0.375	**0.385**
TAC 2008	Baseline 1	0.101	0.217	0.186
	OPM-1	0.102	0.256	0.195
	OPM-2	0.113	0.245	0.208
	GOSM	0.102	0.241	0.216
	PPM	0.103	0.268	**0.231**

Table 2 showed that PPM achieved around 6% and 5% improvement in *F* value compared with Baseline 1 in OpQA and TAC2008, respectively.

5 Related Work

Research on opinion summarization started mostly on review-type data, and much progress has been made in automatic sentiment summarization in the review domain [20]. These summarizations referred to as feature-based summarization or aspect summarizations are extracted from a collection of reviews on some specific product. Benefited from the limited topics and fixed sentiment words in specific domain, technologies such as LDA, LSA, pLSA, have been utilized and they achieved good performance in product review [21, 22, 23, 24]. In this paper, we focus on TOS, which is about general domain summarization, and the above works are out of the scope of TOS due to the constraints of limited targets and fixed sentiment words.

For TOS, lots of work concentrates on term weighting to improve the precision of sentence ranking. A weighted sentiment dictionary was generated from previous Text Retrieval Conference (TREC) relevance data [11]. This dictionary was submitted as a query to a search engine to get an initial query-independent opinion score of all retrieved documents. Similarly, a pseudo opinionated word composed of all opinion words was first created, and then used to estimate the opinion score of a document [3]. This method was shown to be very effective in TREC evaluations.

Ernsting et al. applied the KL divergence to weigh opinionated word [25]. However, the weights of the terms in the sentiment word dictionary were biased towards the terms with high values. Experimental results showed that this method had detrimental effect on the performance. [9] followed the KL divergence measurement and made a positive experimental result by taking term frequency into consideration.

Li et al. proposed a new representation based on word pair [6] for TOI. With the help of word pair, the associative information between the opinion and its corresponding target could be uniformly represented. However, [6] didn't give an explicit approach to weigh word pair but utilize the relationship between word pair and document instead, which is in accordance with "*relevance*".

In this paper, we also utilize word pair to represent TOI. Different from previous work, we present a weighting scheme, which regards both topic-specific words and sentiment words as informative content to represent topic-specific information and opinionated information. Moreover, regarding the specialty of TOS, we propose a method to estimate individual associative score for each word pair to measure the association of topic-specific information and opinionated information, and take *negation* into consideration and integrate it into word pair for TOS.

6 Conclusion and Future Work

In this work, we present a method for topic-specific opinion summarization inspired by the representation of word pair. Based on word pair, we further propose a weighting scheme so that both topic-specific words and sentiment words are weighed. We also provide a method by normalizing PMI between sentiment word and topic-specific word to compute the associative score for individual word pair. Thus, the topic-specific opinionated information of individual opinion expression is able to be well expressed and measured. We integrate word pair into the PageRank model for sentence ranking and adopt maximal marginal relevance method to extract salient sentences as the result of TOS.

In the future, more research is required in the following directions:

(1) Deeper NLP techniques e.g., discourse analysis [26], dependency parser, collocation identification[17] may help to extract word pair and understand the meaning of opinion so as to improve the accuracy.
(2) Opinion holder is another important attribute of TOI [27]. It would be interesting to study opinion holders for QOS.
(3) Since the new weighting scheme and the trade-off parameter indicate topic-specific opinionated information effectively, it is worth further study on other opinion oriented applications.

Acknowledgments. This work is partially supported by National 863 program of China (Grant No. 2009AA01Z150), the Innovation and Technology Fund of Hong Kong SAR (Project No. GHP/036/09SZ) and 2010/11 CUHK Direct Grants (Project No. EE09743).

References

[1] Hu, M., Liu, B.: Mining and summarizing customer reviews. In: Proceedings of the Tenth ACM SIGKDD 2004 (2004)
[2] Huang, X., Croft, W.: A Unified Relevance Model for Opinion Retrieval. In: Proceedings of CIKM (2009)
[3] Na, S.-H., Lee, Y., Nam, S.-H., Lee, J.-H.: Improving Opinion Retrieval Based on Query-Specific Sentiment Lexicon. In: Boughanem, M., Berrut, C., Mothe, J., Soule-Dupuy, C. (eds.) ECIR 2009. LNCS, vol. 5478, pp. 734–738. Springer, Heidelberg (2009)
[4] Zhang, M., Ye, X.: A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval. In: SIGIR 2008 (2008)

[5] Yang, K., Yu, N., Valerio, A., Zhang, H.: WIDIT in TREC-2006 Blog track. In: Proceedings of TREC (2006)
[6] Li, B., Zhou, L., Feng, S., Wong, K.F.: A unified graph model for sentence-based opinion retrieval. In: ACL 2010 (2010)
[7] Dang, H.: Overview of the TAC 2008 opinion question answering and summarization tasks. In: TAC 2008 (2008)
[8] Stoyanov, V., Cardie, C.: Toward opinion summarization: linking the sources. In: Proceedings of the Workshop on Sentiment and Subjectivity in Text (2006)
[9] He, B., Macdonald, C., He, J., Ounis, I.H.: An effective statistical approach to blog post opinion retrieval. In: CIKM 2008 (2008)
[10] Turney, P.: Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In: ACL 2002 (2002)
[11] Amati, G.: Probabilistic models for information retrieval based on Divergence from Randomness. PhD thesis, University of Glasgow (2003)
[12] Kim, J., Li, J., Lee, J.: Discovering the discriminative views: Measuring term weights for sentiment analysis. In: ACL-IJCNLP 2009 (2009)
[13] Esuli, A., Sebastiani, F.: Sentiwordnet: A publicly available lexical resource for opinion mining. In: LREC (2006)
[14] Erkan, G., Radev, D.R.: Lexpagerank: Prestige in multi-document text summarization. In: EMNLP 2004 (2004)
[15] Macdonald, C., Ounis, I.: Overview of the TREC-2006 Blog Track. In: TREC 2006 (2006)
[16] Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: EMNLP 2005 (2005)
[17] Xia, Y., Xu, R., Wong, K.F., Feng: The unified collocation framework for opinion mining. In: International Conference on Machine Learning and Cybernetics 2007 (2007)
[18] Varma, V., Pingali, P., Katragadda, R., Krishna, S., Ganesh, S., Sarvabhotla, K., Garapati, H., Gopisetty, H., Reddy, V.B., Reddy, K., Bysani, P., Bharadwaj, R.: IIIT Hyderabad at TAC 2008. In: TAC 2008 (2008)
[19] Li, F., Tang, Y., Huang, M., Zhu, X.: Answering opinion questions with random walks on graphs. In: ACL 2009 (2009)
[20] Lerman, K., Godensohn, S., McDonald, R.: Sentiment Summarization: Evaluating and learning User Preferences. In: EACL 2009 (2009)
[21] Paul, M., Zhai, C., Girju, R.: Summarizing contrastive viewpoints in opinionated text. In: EMNLP 2010 (2010)
[22] Zhao, X., Jiang, J., Yan, H., Li, X.: Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid. In: EMNLP 2010 (2010)
[23] Lin, C., He, Y., Everson, R.: A comparative study of Bayesian models for unsupervised sentiment detection. In: ACL 2010 (2010)
[24] Blitzer, J., Dredze, M., Pereira, F.: Biographies, bollywood, boomboxes and blenders: Domain adaptationfor sentiment classification. In: ACL (2007)
[25] Ernsting, B., Weerkamp, W., de Rijke, M.: Language modeling approaches to blog post and feed finding. In: TREC 2007 (2007)
[26] Zhou, L., Li, B., Gao, W., Wei, Z., Wong, K.F.: Unsupervised Discovery of Discourse Relations for Eliminating Intra-sentence Polarity Ambiguities. In: Proceedings of EMNLP 2011 (Oral presentation), Edinburgh, Scotland, July 27-31 (2011)
[27] Xu, R., Wong, K.F., Xia, Y.: Opinmine - Opinion Analysis System by CUHK for NTCIR-6 Pilot Task. In: Proceedings of NTCIR-6