

# A Unified Graph Model for Sentence-based Opinion Retrieval

**Binyang Li, Lanjun Zhou, Shi Feng, Kam-Fai Wong**  
Department of Systems Engineering and Engineering Management  
The Chinese University of Hong Kong  
{byli, ljzhou, sfeng, kfwong}@se.cuhk.edu.hk

## Abstract

There is a growing research interest in opinion retrieval as on-line users' opinions are becoming more and more popular in business, social networks, etc. Practically speaking, the goal of opinion retrieval is to retrieve documents, which entail opinions or comments, relevant to a target subject specified by the user's query. A fundamental challenge in opinion retrieval is information representation. Existing research focuses on document-based approaches and documents are represented by *bag-of-word*. However, due to loss of contextual information, this representation fails to capture the associative information between an opinion and its corresponding target. It cannot distinguish different degrees of a sentiment word when associated with different targets. This in turn seriously affects opinion retrieval performance. In this paper, we propose a sentence-based approach based on a new information representation, namely topic-sentiment word pair, to capture intra-sentence contextual information between an opinion and its target. Additionally, we consider inter-sentence information to capture the relationships among the opinions on the same topic. Finally, the two types of information are combined in a unified graph-based model, which can effectively rank the documents. Compared with existing approaches, experimental results on the COAE08 dataset showed that our graph-based model achieved significant improvement.

## 1 Introduction

In recent years, there is a growing interest in sharing personal opinions on the Web, such as product reviews, economic analysis, political polls, etc. These opinions cannot only help independent users make decisions, but also obtain valuable feedbacks (Pang et al., 2008). Opinion oriented research, including sentiment classifica-

tion, opinion extraction, opinion question answering, and opinion summarization, etc. are receiving growing attention (Wilson, et al., 2005; Liu et al., 2005; Oard et al., 2006). However, most existing works concentrate on analyzing opinions expressed in the documents, and none on how to represent the information needs required to retrieve opinionated documents. In this paper, we focus on opinion retrieval, whose goal is to find a set of documents containing not only the query keyword(s) but also the relevant opinions. This requirement brings about the challenge on how to represent information needs for effective opinion retrieval.

In order to solve the above problem, previous work adopts a 2-stage approach. In the first stage, relevant documents are determined and ranked by a score, i.e. *tf-idf* value. In the second stage, an opinion score is generated for each relevant document (Macdonald and Ounis, 2007; Oard et al., 2006). The opinion score can be acquired by either machine learning-based sentiment classifiers, such as SVM (Zhang and Yu, 2007), or a sentiment lexicons with weighted scores from training documents (Amati et al., 2007; Hannah et al., 2007; Na et al., 2009). Finally, an overall score combining the two is computed by using a score function, e.g. linear combination, to re-rank the retrieved documents.

Retrieval in the 2-stage approach is based on document and document is represented by *bag-of-word*. This representation, however, can only ensure that there is at least one opinion in each relevant document, but it cannot determine the relevance pairing of individual opinion to its target. In general, by simply representing a document in *bag-of-word*, contextual information i.e. the corresponding target of an opinion, is neglected. This may result in possible mismatch between an opinion and a target and in turn affects opinion retrieval performance. By the same token, the effect to documents consisting of mul-

multiple topics, which is common in blogs and on-line reviews, is also significant. In this setting, even if a document is regarded as opinionated, it cannot ensure that all opinions in the document are indeed relevant to the target concerned. Therefore, we argue that existing information representation i.e. *bag-of-word*, cannot satisfy the information needs for opinion retrieval.

In this paper, we propose to handle opinion retrieval in the granularity of sentence. It is observed that a complete opinion is always expressed in one sentence, and the relevant target of the opinion is mostly the one found in it. Therefore, it is crucial to maintain the associative information between an opinion and its target within a sentence. We define the notion of a topic-sentiment word pair, which is composed of a topic term (i.e. the target) and a sentiment word (i.e. opinion) of a sentence. Word pairs can maintain intra-sentence contextual information to express the potential relevant opinions. In addition, inter-sentence contextual information is also captured by word pairs to represent the relationship among opinions on the same topic. In practice, the inter-sentence information reflects the degree of a word pair. Finally, we combine both intra-sentence and inter-sentence contextual information to construct a unified undirected graph to achieve effective opinion retrieval.

The rest of the paper is organized as follows. In Section 2, we describe the motivation of our approach. Section 3 presents a novel unified graph-based model for opinion retrieval. We evaluated our model and the results are presented in Section 4. We review related works on opinion retrieval in Section 5. Finally, in Section 6, the paper is concluded and future work is suggested.

## 2 Motivation

In this section, we start from briefly describing the objective of opinion retrieval. We then illustrate the limitations of current opinion retrieval approaches, and analyze the motivation of our method.

### 2.1 Formal Description of Problem

Opinion retrieval was first presented in the TREC 2006 Blog track, and the objective is to retrieve documents that express an opinion about a given target. The opinion target can be a “traditional” named entity (e.g. a name of person, location, or organization, etc.), a concept (e.g. a type of technology), or an event (e.g. presidential

election). The topic of the document is not required to be the same as the target, but an opinion about the target has to be presented in the document or one of the comments to the document (Macdonald and Ounis, 2006). Therefore, in this paper we regard the information needs for opinion retrieval as *relevant opinion*.

### 2.2 Motivation of Our Approach

In traditional information retrieval (IR) *bag-of-word* representation is the most common way to express information needs. However, in opinion retrieval, information need target at *relevant opinion*, and this renders *bag-of-word* representation ineffective.

Consider the example in Figure 1. There are three sentences *A*, *B*, and *C* in a document  $d_i$ . Now given an opinion-oriented query *Q* related to ‘Avatar’. According to the conventional 2-stage opinion retrieval approach,  $d_i$  is represented by a *bag-of-word*. Among the words, there is a topic term *Avatar* ( $t_1$ ) occurring twice, i.e. *Avatar in A* and *Avatar in C*, and two sentiment words *comfortable* ( $o_1$ ) and *favorite* ( $o_2$ ) (refer to Figure 2 (a)). In order to rank this document, an overall score of the document  $d_i$  is computed by a simple combination of the relevant score ( $Score_{rel}$ ) and the opinion score ( $Score_{op}$ ), e.g. equal weighted linear combination, as follows.

$$Score_{doc} = Score_{rel} + Score_{op}$$

For simplicity, we let  $Score_{rel} = tf_Q \times idf_Q$ , and  $Score_{op}$  be computed by using lexicon-based method:  $Score_{op} = weight_{comfortable} + weight_{favorite}$ .

- A. 阿凡达明日将在中国上映。  
Tomorrow, Avatar will be shown in China.
- B. 我预订到了IMAX影院中最舒服的位子。  
I've reserved a comfortable seat in IMAX.
- C. 阿凡达是我最喜欢的一部3D电影。  
Avatar is my favorite 3D movie.

Figure 1: A retrieved document  $d_i$  on the target ‘Avatar’.

Although *bag-of-word* representation achieves good performance in retrieving relevant documents, our study shows that it cannot satisfy the information needs for retrieval of *relevant opinion*. It suffers from the following limitations:

(1) It cannot maintain contextual information; thus, an opinion may not be related to the target of the retrieved document is neglected. In this example, only the opinion *favorite* ( $o_2$ ) on *Avatar in C* is the *relevant opinion*. But due to loss of contextual information between the opinion and its corresponding target, *Avatar in A* and *com-*

comfortable ( $o_1$ ) are also regarded as *relevant opinion* mistakenly, creating a false positive. In reality *comfortable* ( $o_1$ ) describes “the seats in IMAX”, which is an irrelevant opinion, and sentence  $A$  is a factual statement rather than an opinion statement.

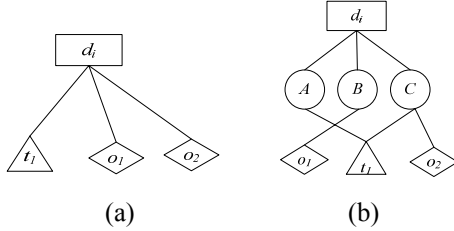


Figure 2: Two kinds of information representation of opinion retrieval. ( $t_1$ =‘Avatar’  $o_1$ = ‘comfortable’,  $o_2$ =‘favorite’)

(1) Current approaches cannot capture the relationship among opinions about the same topic. Suppose there is another document including sentence  $C$  which expresses the same opinion on *Avatar*. Existing information representation simply does not cater for the two identical opinions from different documents. In addition, if many documents contain opinions on *Avatar*, the relationship among them is not clearly represented by existing approaches.

In this paper, we process opinion retrieval in the granularity of sentence as we observe that a complete opinion always exists within a sentence (refer to Figure 2 (b)). To represent a *relevant opinion*, we define the notion of topic-sentiment word pair, which consists of a topic term and a sentiment word. A word pair maintains the associative information between the two words, and enables systems to draw up the relationship among all the sentences with the same opinion on an identical target. This relationship information can identify all documents with sentences including the sentiment words and to determine the contributions of such words to the target (topic term). Furthermore, based on word pairs, we designed a unified graph-based method for opinion retrieval (see later in Section 3).

### 3 Graph-based model

#### 3.1 Basic Idea

Different from existing approaches which simply make use of document relevance to reflect the relevance of opinions embedded in them, our approach concerns more on identifying the relevance of individual opinions. Intuitively, we believed that the more *relevant opinions* appear in a document, the more relevant is that document for subsequent opinion analysis operations.

Further, since the lexical scope of an opinion does not usually go beyond a sentence, we propose to handle opinion retrieval in the granularity of sentence.

Without loss of generality, we assume that there is a document set  $\mathcal{D} = \{d_1, d_2, d_3, \dots, d_n\}$ , and a specific query  $\mathcal{Q} = \{q_1, q_2, q_3, \dots, q_z\}$ , where  $q_1, q_2, q_3, \dots, q_z$  are query keywords. Opinion retrieval aims at retrieving documents from  $\mathcal{D}$  with *relevant opinion* about the query  $\mathcal{Q}$ . In addition, we construct a sentiment word lexicon  $V_o$  and a topic term lexicon  $V_t$  (see Section 4). To maintain the associative information between the target and the opinion, we consider the document set as a *bag of sentences*, and define a sentence set as  $\mathcal{S} = \{s_1, s_2, s_3, \dots, s_N\}$ . For each sentence, we capture the intra-sentence information through the topic-sentiment word pair.

*Definition 1.* topic-sentiment word pair  $p_{ij}$  consists of two elements, one is from  $V_t$ , and the other one is from  $V_o$ .

$$p_{ij} = \{ \langle t_i, o_j \rangle \mid t_i \in V_t, o_j \in V_o \}.$$

The topic term from  $V_t$  determines relevance by the query term matching, and the sentiment word from  $V_o$  is used to express an opinion. We use the word pair to maintain the associative information between the topic term and the opinion word (also referred to as sentiment word). The word pair is used to identify a *relevant opinion* in a sentence. In Figure 2 (b),  $t_1$ , i.e. *Avatar* in  $C$ , is a topic term relevant to the query, and  $o_2$  (‘favorite’) is supposed to be an opinion; and the word pair  $\langle t_1, o_2 \rangle$  indicates sentence  $C$  contains a *relevant opinion*. Similarly, we map each sentence in word pairs by the following rule, and express the intra-sentence information using word pairs.

For each sentiment word of a sentence, we choose the topic term with minimum distance as the other element of the word pair:

$$s_l \rightarrow \{ \langle t_i, o_j \rangle \mid t_i = \min \text{Dist}(t_i, o_j) \text{ for each } o_j \}$$

According to the mapping rule, although a sentence may give rise to a number of word pairs, only the pair with the minimum word distance is selected. We do not take into consideration of the other words in a sentence as *relevant opinions* are generally formed in close proximity. A sentence is regarded non-opinionated unless it contains at least one word pair.

In practice, not all word pairs carry equal weights to express a *relevant opinion* as the contribution of an opinion word differs from different target topics, and vice versa. For example, the word pair  $\langle t_1, o_2 \rangle$  should be more probable as a *relevant opinion* than  $\langle t_1, o_1 \rangle$ . To consider

that, inter-sentence contextual information is explored. This is achieved by assigning a weight to each word pair to measure their associative degrees to different queries. We believe that the more a word pair appears the higher should be the weight between the opinion and the target in the context.

We will describe how to utilize intra-sentence contextual information to express *relevant opinion*, and inter-sentence information to measure the degree of each word pair through a graph-based model in the following section.

### 3.2 HITS Model

We propose an opinion retrieval model based on HITS, a popular graph ranking algorithm (Kleinberg, 1999). By considering both intra-sentence information and inter-sentence information, we can determine the weight of a word pair and rank the documents.

HITS algorithm distinguishes hubs and authorities in objects. A hub object has links to many authorities. An authority object, which has high-quality content, would have many hubs linking to it. The hub scores and authority scores are computed in an iterative way. Our proposed opinion retrieval model contains two layers. The upper level contains all the topic-sentiment word pairs  $p_{ij} = \{ \langle t_i, o_j \rangle \mid t_i \in V_t, o_j \in V_o \}$ . The lower level contains all the documents to be retrieved. Figure 3 gives the bipartite graph representation of the HITS model.

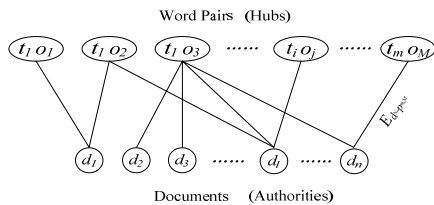


Figure 3: Bipartite link graph.

For our purpose, the word pairs layer is considered as hubs and the documents layer authorities. If a word pair occurs in one sentence of a document, there will be an edge between them. In Figure 3, we can see that the word pair that has links to many documents can be assigned a high weight to denote a strong associative degree between the topic term and a sentiment word, and it likely expresses a *relevant opinion*. On the other hand, if a document has links to many word pairs, the document is with many *relevant opinions*, and it will result in high ranking.

Formally, the representation for the bipartite graph is denoted as  $G = \langle H_p, A_d, E_{dp} \rangle$ , where  $H_p = \{p_{ij}\}$  is the set of all pairs of topic words

and sentiment words, which appear in one sentence.  $A_d = \{d_k\}$  is the set of documents.  $E_{dp} = \{e_{ij}^k \mid p_{ij} \in H_p, d_k \in A_d\}$  corresponds to the connection between documents and topic-sentiment word pairs. Each edge  $e_{ij}^k$  is associated with a weight  $w_{ij}^k \in [0,1]$  denoting the contribution of  $p_{ij}$  to the document  $d_k$ . The weight  $w_{ij}^k$  is computed by the contribution of word pair  $p_{ij}$  in all sentences of  $d_k$  as follows:

$$w_{ij}^k = \frac{1}{|d_k|} \sum_{p_{ij} \in s_l \in d_k} [\lambda \cdot rel(t_i, s_l) + (1 - \lambda) \cdot opn(o_j, s_l)] \quad (1)$$

- $|d_k|$  is the number of sentences in  $d_k$ ;
- $\lambda$  is introduced as the trade-off parameter to balance the  $rel(t_i, s_l)$  and  $opn(o_j, s_l)$ ;
- $rel(t_i, s_l)$  is computed to judge the relevance of  $t_i$  in  $s_l$  which belongs to  $d_k$ ;

$$rel(t_i, s_l) = tf_{t_i, s_l} \times isf_{t_i} \quad (2)$$

where  $tf_{t_i, s_l}$  is the number of  $t_i$  appears in  $s_l$ , and

$$isf_{t_i} = \log\left(\frac{N+1}{0.5+sf_{t_i}}\right) \quad (3)$$

where  $sf_{t_i}$  is the number of sentences that the word  $t_i$  appears in.

- $opn(o_j, s_l)$  is the contribution of  $o_j$  in  $s_l$  which belongs to  $d_k$ .

$$opn(o_j, s_l) = \frac{tf_{o_j, s_l}}{tf_{o_j, s_l} + 0.5 + (1.5 \times \frac{len(s_l)}{asl})} \quad (4)$$

where  $asl$  is the average number of sentences in  $d_k$ ;  $tf_{o_j, s_l}$  is the number of  $o_j$  appears in  $s_l$  (Allan et al., 2003; Otterbacher et al., 2005).

It is found that the contribution of a sentiment word  $o_j$  will not decrease even if it appears in all the sentences. Therefore in Equation 4, we just use the length of a sentence instead of  $isf_{o_j}$  to normalize long sentences which would likely contain more sentiment words.

The authority score  $AuthScore^{(T+1)}(d_k)$  of document  $d_k$  and a hub score  $HubScore^{(T+1)}(p_{ij})$  of  $p_{ij}$  at the  $(T+1)^{th}$  iteration are computed based on the hub scores and authority scores in the  $T^{th}$  iteration as follows.

$$AuthScore^{(T+1)}(d_k) = \sum_{p_{ij} \in H_p} w_{ij}^k \times HubScore^T(p_{ij}) \quad (5)$$

$$HubScore^{(T+1)}(p_{ij}) = \sum_{d_k \in A_d} w_{ij}^k \times AuthScore^T(d_k) \quad (6)$$

We let  $L = (L_{i,j})_{|H_p| \times |A_d|}$  denote the adjacency matrix.

$$\vec{a}^{(T+1)} = L \vec{h}^{(T)} \quad (7)$$

$$\vec{h}^{(T+1)} = L^T \vec{a}^{(T)} \quad (8)$$

where  $\vec{a}^{(T)} = [AuthScore^T(d_k)]_{|A_d| \times 1}$  is the vector of authority scores for documents at the  $T^{th}$  iteration and  $\vec{h}^{(T)} = [HubScore^T(p_{ij})]_{|H_p| \times 1}$  is the vector of hub scores for the word pairs at  $T^{th}$  iteration. In order to ensure convergence of the iterative form,  $\vec{a}$  and  $\vec{h}$  are normalized in each iteration cycle.

For computation of the final scores, the initial scores of all documents are set to  $\frac{1}{\sqrt{n}}$ , and topic-sentiment word pairs are set to  $\frac{1}{\sqrt{m \times M}}$ . The above iterative steps are then used to compute the new scores until convergence. Usually the convergence of the iteration algorithm is achieved when the difference between the scores computed at two successive iterations for any nodes falls below a given threshold (Wan et al., 2008; Li et al., 2009; Erkan and Radev, 2004). In our model, we use the hub scores to denote the associative degree of each word pair and the authority scores as the total scores. The documents are then ranked based on the total scores.

## 4 Experiment

We performed the experiments on the Chinese benchmark dataset to verify our proposed approach for opinion retrieval. We first tested the effect of the parameter  $\lambda$  of our model. To demonstrate the effectiveness of our opinion retrieval model, we compared its performance with the same of other approaches. In addition, we studied each individual query to investigate the influence of query to our model. Furthermore, we showed the top-5 highest weight word pairs of 5 queries to further demonstrate the effect of word pair.

### 4.1 Experiment Setup

#### 4.1.1 Benchmark Datasets

Our experiments are based on the Chinese benchmark dataset, COAE08 (Zhao et al., 2008). COAE dataset is the benchmark data set for the opinion retrieval track in the Chinese Opinion Analysis Evaluation (COAE) workshop, consisting of blogs and reviews. 20 queries are provided in COAE08. In our experiment, we created relevance judgments through pooling method, where documents are ranked at different levels: irrelevant, relevant but without opinion, and relevant with opinion. Since polarity is not considered, all relevant documents with opinion are classified into the same level.

#### 4.1.2 Sentiment Lexicon

In our experiment, the sentiment lexicon is composed by the following resources (Xu et al., 2007):

- (1) *The Lexicon of Chinese Positive Words*, which consists of 5,054 positive words and the *Lexicon of Chinese Negative Words*, which consists of 3,493 negative words;

- (2) The opinion word lexicon provided by National Taiwan University which consists of 2,812 positive words and 8,276 negative words;
- (3) Sentiment word lexicon and comment word lexicon from *HowNet*. It contains 1836 positive sentiment words, 3,730 positive comments, 1,254 negative sentiment words and 3,116 negative comment words.

The different graphemes corresponding to Traditional Chinese and Simplified Chinese are both considered so that the sentiment lexicons from different sources are applicable to process Simplified Chinese text. The lexicon was manually verified.

#### 4.1.3 Topic Term Collection

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

The dictionary-based method utilizes Wikipedia (Popescu and Etzioni, 2005) to find an entry page for a phrase or a single term in a query. If such an entry exists, all titles of the entry page are extracted as synonyms of the query concept. For example, if we search “绿坝” (Green Tsunami, a firewall) in Wikipedia, it is re-directed to an entry page titled “花季护航” (Youth Escort). This term is then added as a synonym of “绿坝” (Green Tsunami) in the query. Synonyms are treated the same as the original query terms in a retrieval process. The content words in the entry page are ranked by their frequencies in the page. The top- $k$  terms are returned as potential expanded topic terms.

The second query expansion method is a web-based method. It is similar to the pseudo relevance feedback expansion but using web documents as the document collection. The query is submitted to a web search engine, such as Google, which returns a ranked list of documents. In the top- $n$  documents, the top- $m$  topic terms which are highly correlated to the query terms are returned.

### 4.2 Performance Evaluation

#### 4.2.1 Parameter Tuning

We first studied how the parameter  $\lambda$  (see Equation 1) influenced the mean average precision (MAP) in our model. The result is given in Figure 4.

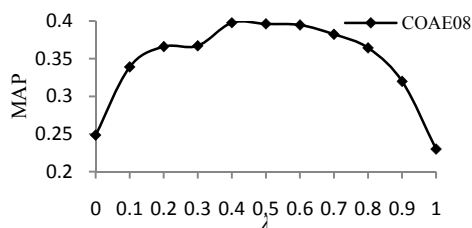


Figure 4: Performance of MAP with varying  $\lambda$ .

Best MAP performance was achieved in COAE08 evaluation, when  $\lambda$  was set between 0.4 and 0.6. Therefore, in the following experiments, we set  $\lambda = 0.4$ .

#### 4.2.2 Opinion Retrieval Model Comparison

To demonstrate the effectiveness of our proposed model, we compared it with the following models using different evaluation metrics:

- (1) IR: We adopted a classical information retrieval model, and further assumed that all retrieved documents contained *relevant opinions*.
- (2) Doc: The 2-stage document-based opinion retrieval model was adopted. The model used sentiment lexicon-based method for opinion identification and a conventional information retrieval method for relevance detection.
- (3) ROSC: This was the model which achieved the best run in TREC Blog 07. It employed machine learning method to identify opinions for each sentence, and to determine the target topic by a *NEAR* operator.
- (4) ROCC: This model was similar to ROSC, but it considered the factor of sentence and regarded the count of relevant opinionated sentence to be the opinion score (Zhang and Yu, 2007). In our experiment, we treated this model as the evaluation baseline.
- (5) GORM: our proposed graph-based opinion retrieval model.

Approach	COAE08			
	Evaluation metrics			
Run id	MAP	R-pre	bPref	P@10
IR	0.2797	0.3545	0.2474	0.4868
Doc	0.3316	0.3690	0.3030	0.6696
ROSC	0.3762	0.4321	0.4162	0.7089
Baseline	0.3774	0.4411	0.4198	0.6931
GORM	<b>0.3978</b>	<b>0.4835</b>	<b>0.4265</b>	<b>0.7309</b>

Table 1: Comparison of different approaches on COAE08 dataset, and the best is highlighted.

Most of the above models were originally designed for opinion retrieval in English, and re-designed them to handle Chinese opinionated documents. We incorporated our own Chinese sentiment lexicon for this purpose. In our experiments, in addition to MAP, other metrics such as R-precision (R-prec), binary Preference (bPref)

and Precision at 10 documents (P@10) were also used. The evaluation results based on these metrics are shown in Table 1.

Table 1 summarized the results obtained. We found that GORM achieved the best performance in all the evaluation metrics. Our baseline, ROSC and GORM which were sentence-based approaches achieved better performance than the document-based approaches by 20% in average. Moreover, our GORM approach did not use machine learning techniques, but it could still achieve outstanding performance.

To study GORM influenced by different queries, the MAP from median average precision on individual topic was shown in Figure 5.

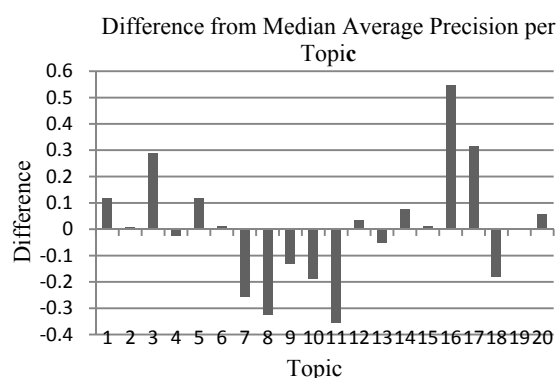


Figure 5: Difference of MAP from Median on COAE08 dataset. (MAP of Median is 0.3724)

As shown in Figure 5, the MAP performance was very low on topic 8 and topic 11. Topic 8, i.e. ‘成龙’ (Jackie Chan), it was influenced by topic 7, i.e. ‘李连杰’ (Jet Lee) as there were a number of similar relevant targets for the two topics, and therefore many word pairs ended up the same. As a result, documents belonging to topic 7 and topic 8 could not be differentiated, and they both performed badly. In order to solve this problem, we extracted the topic term with highest relevant weight in the sentence to form word pairs so that it reduce the impact on the topic terms in common. 24% and 30% improvement were achieved, respectively.

As to topic 11, i.e. ‘指环王’ (Lord of King), there were only 8 relevant documents without any opinion and 14 documents with *relevant opinions*. As a result, the graph constructed by insufficient documents worked ineffectively.

Except for the above queries, GORM performed well in most of the others. To further investigate the effect of word pair, we summarized the top-5 word pairs with highest weight of 5 queries in Table 2.

Top-5 MAP				
陈凯歌 Chen Kaige	国六条 Six States	宏观调控 Macro-regulation	周星驰 Stephen Chow	Vista Vista
<陈凯歌 支持> Chen Kaige Support	<房价 上涨> Room rate Rise	<经济 平稳> Economics Steady	<电影 喜欢> Movie Like	<价格 贵> Price Expensive
<陈凯歌 最佳> Chen Kaige Best	<调控 加强> Regulate Strengthen	<价格 上涨> Price Rise	<周星驰 喜欢> Stephen Chow Like	<微软 喜欢> Microsoft Like
<《无极》 骂> Limitless Revile	<中央 加强> CCP Strengthen	<发展 平稳> Development Steady	<主角 最佳> Protagonist Best	<Vista 推荐> Vista Recommend
<影片 优秀> Movie Excellent	<房价 平稳> Room rate Steady	<消费 上涨> Consume Rise	<喜剧 好> Comedy Good	<问题 重要> Problem Vital
<阵容 强大的> Cast Strong	<住房 保障> Housing Security	<社会 保障> Social Security	<作品 精彩> Works Splendid	<性能 不> Performance No

Table 2: Top-5 highest weight word pairs for 5 queries in COAE08 dataset.

Table 2 showed that most word pairs could represent the *relevant opinions* about the corresponding queries. This showed that inter-sentence information was very helpful to identify the associative degree of a word pair. Furthermore, since word pairs can indicate relevant opinions effectively, it is worth further study on how they could be applied to other opinion oriented applications, e.g. opinion summarization, opinion prediction, etc.

## 5 Related Work

Our research focuses on *relevant opinion* rather than on relevant document retrieval. We, therefore, review related works in opinion identification research. Furthermore, we do not support the conventional 2-stage opinion retrieval approach. We conducted literature review on unified opinion retrieval models and related work in this area is presented in the section.

### 5.1 Lexicon-based Opinion Identification

Different from traditional IR, opinion retrieval focuses on the opinion nature of documents. During the last three years, NTICR and TREC evaluations have shown that sentiment lexicon-based methods led to good performance in opinion identification.

A lightweight lexicon-based statistical approach was proposed by Hannah et al. (2007). In this method, the distribution of terms in relevant opinionated documents was compared to their distribution in relevant fact-based documents to calculate an opinion weight. These weights were used to compute opinion scores for each retrieved document. A weighted dictionary was generated from previous TREC relevance data (Amati et al., 2007). 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 (Na et al., 2009). This method was shown to be very effective in TREC evaluations (Lee et al., 2008). More recently, Huang and Croft (2009) proposed an effective relevance model, which integrated both query-independent and query-dependent sentiment words into a mixture model.

In our approach, we also adopt sentiment lexicon-based method for opinion identification. Unlike the above methods, we generate a weight to a sentiment word for each target (associated topic term) rather than assign a unified weight or an equal weight to the sentiment word for the whole topics. Besides, in our model no training data is required. We just utilize the structure of our graph to generate a weight to reflect the associative degree between the two elements of a word pair in different context.

### 5.2 Unified Opinion Retrieval Model

In addition to conventional 2-stage approach, there has been some research on unified opinion retrieval models.

Eguchi and Lavrenko proposed an opinion retrieval model in the framework of generative language modeling (Eguchi and Lavrenko, 2006). They modeled a collection of natural language documents or statements, each of which consisted of some topic-bearing and some sentiment-bearing words. The sentiment was either represented by a group of predefined seed words, or extracted from a training sentiment corpus. This model was shown to be effective on the MPQA corpus.

Mei et al. tried to build a fine-grained opinion retrieval system for consumer products (Mei et al., 2007). The opinion score for a product was a mixture of several facets. Due to the difficulty in

associating sentiment with products and facets, the system was only tested using small scale text collections.

Zhang and Ye proposed a generative model to unify topic relevance and opinion generation (Zhang and Ye, 2008). This model led to satisfactory performance, but an intensive computation load was inevitable during retrieval, since for each possible candidate document, an opinion score was summed up from the generative probability of thousands of sentiment words.

Huang and Croft proposed a unified opinion retrieval model according to the Kullback-Leibler divergence between the two probability distributions of opinion relevance model and document model (Huang and Croft, 2009). They divided the sentiment words into query-dependent and query-independent by utilizing several sentiment expansion techniques, and integrated them into a mixed model. However, in this model, the contribution of a sentiment word was its corresponding incremental mean average precision value. This method required that large amount of training data and manual labeling.

Different from the above opinion retrieval approaches, our proposed graph-based model processes opinion retrieval in the granularity of sentence. Instead of *bag-of-word*, the sentence is split into word pairs which can maintain the contextual information. On the one hand, word pair can identify the *relevant opinion* according to intra-sentence contextual information. On the other hand, it can measure the degree of a *relevant opinion* by considering the inter-sentence contextual information.

## 6 Conclusion and Future Work

In this work we focus on the problem of opinion retrieval. Different from existing approaches, which regard document relevance as the key indicator of opinion relevance, we propose to explore the relevance of individual opinion. To do that, opinion retrieval is performed in the granularity of sentence. We define the notion of word pair, which can not only maintain the association between the opinion and the corresponding target in the sentence, but it can also build up the relationship among sentences through the same word pair. Furthermore, we convert the relationships between word pairs and sentences into a unified graph, and use the HITS algorithm to achieve document ranking for opinion retrieval. Finally, we compare our approach with existing methods.

Experimental results show that our proposed model performs well on COAE08 dataset.

The novelty of our work lies in using word pairs to represent the information needs for opinion retrieval. On the one hand, word pairs can identify the *relevant opinion* according to intra-sentence contextual information. On the other hand, word pairs can measure the degree of a *relevant opinion* by taking inter-sentence contextual information into consideration. With the help of word pairs, the information needs for opinion retrieval can be represented appropriately.

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

- (1) Since word pairs can indicate relevant opinions effectively, it is worth further study on how they could be applied to other opinion oriented applications, e.g. opinion summarization, opinion prediction, etc.
- (2) The characteristics of blogs will be taken into consideration, i.e., the post time, which could be helpful to create a more time sensitivity graph to filter out fake opinions.
- (3) Opinion holder is another important role of an opinion, and the identification of opinion holder is a main task in NTCIR. It would be interesting to study opinion holders, e.g. its seniority, for opinion retrieval.

**Acknowledgements:** This work is partially supported by the Innovation and Technology Fund of Hong Kong SAR (No. ITS/182/08) and National 863 program (No. 2009AA01Z150). Special thanks to Xu Hongbo for providing the Chinese sentiment resources. We also thank Bo Chen, Wei Gao, Xu Han and anonymous reviewers for their helpful comments.

## References

- James Allan, Courtney Wade, and Alvaro Bolivar. 2003. Retrieval and novelty detection at the sentence level. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 314-321. ACM.
- Giambattista Amati, Edgardo Ambrosi, Marco Bianchi, Carlo Gaibisso, and Giorgio Gambosi. 2007. FUB, IASI-CNR and University of Tor Vergata at TREC 2007 Blog Track. In *Proceedings of the 15<sup>th</sup> Text Retrieval Conference*.
- Koji Eguchi and Victor Lavrenko. Sentiment retrieval using generative models. 2006. In *EMNLP '06, Proceedings of 2006 Conference on Empirical Methods in Natural Language Processing*, page 345-354.



- Gunes Erkan and Dragomir R. Radev. 2004. Lexpagerank: Prestige in multi-document text summarization. In *EMNLP '04, Proceedings of 2004 Conference on Empirical Methods in Natural Language Processing*.
- David Hannah, Craig Macdonald, Jie Peng, Ben He, and Iadh Ounis. 2007. University of Glasgow at TREC 2007: Experiments in Blog and Enterprise Tracks with Terrier. In *Proceedings of the 15<sup>th</sup> Text Retrieval Conference*.
- Xuanjing Huang, William Bruce Croft. 2009. A Unified Relevance Model for Opinion Retrieval. In *Proceedings of CIKM*.
- Jon M. Kleinberg. 1999. Authoritative sources in a hyperlinked environment. *J. ACM*, 46(5): 604-632.
- Yeha Lee, Seung-Hoon Na, Jungi Kim, Sang-Hyob Nam, Hun-young Jung, Jong-Hyeok Lee. 2008. KLE at TREC 2008 Blog Track: Blog Post and Feed Retrieval. In *Proceedings of the 15<sup>th</sup> Text Retrieval Conference*.
- Fangtao Li, Yang Tang, Minlie Huang, and Xiaoyan Zhu. 2009. Answering Opinion Questions with Random Walks on Graphs. In *ACL '09, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*.
- Bing Liu, Mingqing Hu, and Junsheng Cheng. 2005. Opinion observer: Analyzing and comparing opinions on the web. In *WWW '05: Proceedings of the 14<sup>th</sup> International Conference on World Wide Web*.
- Craig Macdonald and Iadh Ounis. 2007. Overview of the TREC-2007 Blog Track. In *Proceedings of the 15<sup>th</sup> Text Retrieval Conference*.
- Craig Macdonald and Iadh Ounis. 2006. Overview of the TREC-2006 Blog Track. In *Proceedings of the 14<sup>th</sup> Text Retrieval Conference*.
- Qiaozhu Mei, Xu Ling, Matthew Wondra, Hang Su, and Chengxiang Zhai. 2007. Topic sentiment mixture: Modeling facets and opinions in weblogs. In *WWW '07: Proceedings of the 16 International Conference on World Wide Web*.
- Seung-Hoon Na, Yeha Lee, Sang-Hyob Nam, and Jong-Hyeok Lee. 2009. Improving opinion retrieval based on query-specific sentiment lexicon. In *ECIR '09: Proceedings of the 31<sup>st</sup> annual European Conference on Information Retrieval*, pages 734-738.
- Douglas Oard, Tamer Elsayed, Jianqiang Wang, Yejun Wu, Pengyi Zhang, Eileen Abels, Jimmy Lin, and Dagbert Soergel. 2006. TREC-2006 at Maryland: Blog, Enterprise, Legal and QA Tracks. In *Proceedings of the 15<sup>th</sup> Text Retrieval Conference*.
- Jahna Otterbacher, Gunes Erkan, and Dragomir R. Radev. 2005. Using random walks for question-focused sentence retrieval. In *EMNLP '05, Proceedings of 2005 Conference on Empirical Methods in Natural Language Processing*.
- Larry Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1998. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford University.
- Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2): 1-135.
- Ana-Maria Popescu and Oren Etzioni. 2005. Extracting product features and opinions from reviews. In *EMNLP '05, Proceedings of 2005 Conference on Empirical Methods in Natural Language Processing*.
- Xiaojun Wan and Jianwu Yang. 2008. Multi-document summarization using cluster-based link analysis. In *SIGIR '08: Proceedings of the 31th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 299-306. ACM.
- Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *EMNLP '05, Proceedings of 2005 Conference on Empirical Methods in Natural Language Processing*.
- Ruifeng Xu, Kam-Fai Wong and Yunqing Xia. 2007. Opinmine - Opinion Analysis System by CUHK for NTCIR-6 Pilot Task. In *Proceedings of NTCIR-6*.
- Min Zhang and Xingyao Ye. 2008. A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval. In *SIGIR '08: Proceedings of the 31<sup>st</sup> Annual International ACM SIGIR conference on Research and Development in Information Retrieval*, pages 411-418. ACM.
- Wei Zhang and Clement Yu. 2007. UIC at TREC 2007 Blog Track. In *Proceedings of the 15<sup>th</sup> Text Retrieval Conference*.
- Jun Zhao, Hongbo Xu, Xuanjing Huang, Songbo Tan, Kang Liu, and Qi Zhang. 2008. Overview of Chinese Opinion Analysis Evaluation 2008. In *Proceedings of the First Chinese Opinion Analysis Evaluation*.