Coarse-Fine Opinion Mining – WIA in NTCIR-7 MOAT Task

Ruifeng Xu, Kam-Fai Wong
Department of Systems Engineering and Engineering Management,
The Chinese University of Hong Kong, Shatin, Hong Kong
**Ruifeng Xu is now working in Department of Chinese, Translation and Linguistics,*
City University of Hong Kong, Kowloon, Hong Kong

Yunqing Xia
Center for Speech and Language Technologies, RIIT,
Tsinghua University, Beijing, China

Abstract

This paper presents an opinion analysis system developed by CUHK_PolyU_Tsinghua Web Information Analysis Group (WIA), namely WIA-Opinmine, for NTCIR-7 MOAT Task. Different from most existing opinion mining systems, which recognize opinionated sentences as one-step classification procedure, WIA-Opinmine adopts a multi-pass coarse-fine analysis strategy. A base classifier firstly coarsely estimates the opinion of sentences and the document. The obtained document-level and sentence-level opinions are then incorporated in a complex classifier to re-analyze the opinion of sentences to obtain refined sentence and document opinions. The updated opinion features are feed back to the complex classifier to further refine the opinion analysis. Such circles terminate until the analysis results converge. Similar strategy is adopted in sentence-topic relevance estimation. Furthermore, the mutual reinforcement between the analysis of sentence relevance and sentence opinion are integrated in one framework in WIA-Opinmine. Evaluations on NTCIR-7 MOAT Traditional Chinese and Simplified Chinese sides show that WIA-Opinmine achieves the best precisions performance in five subtasks and the best F performance in three subtasks including polarity determination, opinion holder recognition and opinion target recognition. This results show that the proposed framework integrating coarse-fine opinion mining strategy and the mutual reinforcement between the analysis of sentence relevance and sentence opinion is promising.

Keywords: *Opinion mining, Coarse-Fine opinion mining, mutual reinforcement*

1. Introduction

Opinion mining aims to identify and analyze the opinions from text since the discovered opinions are useful to many applications. Besides, the opinion mining technique promotes the research in information extraction and knowledge discovery such as automatic summarization and question & answer system [1, 2].

Various techniques are proposed to identify document-level and sentence-level opinions in different domains [3, 4]. These approaches were designed for different purposes and different domains. Thus, their performance are difficult to evaluated and compared. For this reason, NTCIR-6 provided a pilot task to evaluate and compare different approaches for multilingual opinion analysis [5]. Based on this, NTCIR-7 Multilingual Opinion Analysis Task (MOAT) provides one more opportunity to evaluate the opinion mining techniques. NTCIR-7 MOAT defines five subtasks [6]:

1. Determine the relevance between each sentence and the topic.
2. Determine the opinion of each sentence. It is a binary classification, opinionated or not.
3. Determine the polarity of each opinionated sentence. The possible polarity values are positive (POS), negative (NEG), or neutral (NEU).
4. Recognize the opinion holder in the opinionated sentence. The opinion holder is the governor of an opinion. Each opinion expression may have at least one opinion holder.
5. Recognize the opinion target in the opinionated sentence. Each opinion expression may have at least one opinion target.

Notice that the first two categories are mandatory and the other three are optional.

CUHK_PolyU_Tsinghua Web Information Analysis Group (WIA) developed WIA-Opinmine system and participated in NTCIR-7 MOAT on Traditional Chinese and Simplified Chinese sides. This system adopts a new framework, which analyzes sentence opinions and sentence relevance following a mutual-reinforced coarse-fine analysis strategy. Such as framework is different from most existing opinion mining systems, which regard opinionated sentence determination as one-step classification problem. The proposed opinion mining framework is a multi-pass analysis procedure. A base classifier is firstly applied to estimate the opinion of each sentence in the document based on word-level, collocation-level and punctuation-level features. The analysis results of sentence opinions generate the document opinion. Considering the document opinion is helpful to sentence opinion analysis, the document-level and sentence-level opinion features are incorporated in a complex classifier to re-analyze the sentence opinions. The obtained refined opinions of sentences and document are feed back to the complex classifier to further refine the sentence opinion analysis. Such circles terminate until the analysis outputs converge. Similar multiple-pass analyses are conducted to estimate the sentence-topic relevance. Furthermore, considering that in a topic-relevant document, an opinionated sentence always focuses on the main target of the document, which means topic-relevant, the analysis of sentence opinion and sentence relevance shown mutual reinforced. Thus, a framework integrating the mutual reinforced analysis of sentence relevance and sentence opinion is designed. Following this framework, WIA-Opinmine system is implemented. Its performance is evaluated in NTCIR-7 MOAT on Traditional Chinese and Simplified Chinese side. In the subtasks of sentence relevance determination and opinionated sentence determination, WIA-Opinmine ranked 4 and 5 among 7 teams. In the subtasks of polarity determination, opinion holder recognition and opinion target recognition, this system ranked the first. Meanwhile, WIA-Opinmine achieves the best precisions in most subtasks. The achieved promising results support the idea of a mutual-reinforced, multi-pass and coarse-fine opinion mining framework.

The rest of this paper is organized as follows. Section 2 briefly reviews the existing works on opinion mining. Section 3 presents the framework design of mutual-reinforced coarse-fine opinion mining. Section 4 presents the implementation issues of WIA-Opinmine. Section 5 gives the evaluation results and finally, Section 6 concludes this paper.

2. Literature Review

Early opinion mining research focus on the identification and polarity determination of sentiment words. Hatzivassiloglou and Mckoeown predicted semantic orientations of sentiment adjectives by analyzing adjective pairs occurring in the corpus [7]. Turney and Litman, and Kamps investigated different unsupervised techniques to determine the polarity of new sentiment words [8, 9]. Furthermore, [10] showed that automatic detection of gradable adjectives is helpful to opinion mining.

In the past few years, different opinion mining techniques have been proposed to identify document- and sentence-level opinions in different applications domains such as news articles [3], product reviews [11], movie reviews [4] and web blogs [12]. These techniques can be categorized into three approaches: (1) Sentimental knowledge based approach, which utilizes linguistic knowledge on sentiment words and opinion-related heuristic rules as clues for opinion analysis. This approach identifies known sentiment words in a given text and uses the product of the polarities of these sentiment words to recognize the sentence opinion. Opinion-related heuristic rules are applied to improve opinion analysis. Typical systems based on this approach include [13] on English texts and [3, 14] on Chinese. (2) Machine learning based approach to train the machine learning based classifiers using sentiment features, such as sentiment words, word bi-grams, word n-grams, syntactic patterns, punctuations and topic-relevant features, etc. for opinion mining. The supervised and unsupervised learning techniques were used to develop a classifier, which is used to classify input sentences into either opinionated or non-opinionated class. Popular classifiers include Naïve Bayes (NB) [15], Maximum Entropy (ME) [16] and Support Vector Machine (SVM) [17]. (3) Combined approach combines sentiment knowledge, machine learning and a general linguistic framework for opinion analysis, such as opinion-related semantic role labeling by using FrameNet [18].

Existing approaches suffer from four major problems: (1) Many of them fundamentally rely on a sentiment lexicon. However, manual construction of a complied sentiment lexicon is impractical. It is hard to expand and maintain. (2) Many sentiment words are context-sensitive, i.e. they hold different polarities depending on context. The characteristics of this kind of context-dependent sentiment words are not well studied. (3) Features based on linguistic knowledge related to opinion expressions are not adequately studied. (4) The size of annotated opinion corpus is not large enough to support effective supervised machine learning.

3. Framework Design

Most existing opinion mining techniques regard opinionated sentence identification as a classification problem. The linguistic features and statistical-based features in the observing sentence are regarded as distinguish features for the classifier for determine the opinion of the sentences. These techniques ignore the influence of opinions of the document and the neighboring sentences to the opinion analysis of the observing sentence. Intuitively, a sentence in a strong polarity document has higher probability to be the same polarity while a sentence in a factual document tends to be factual too. The observations on NTCIR-6 corpus and NTCIR-7 training corpus verify this idea. Naturally, the document-level and sentence-level opinions should be considered in sentence opinion analysis. Meanwhile, humans normally understand the opinion trend of a document coarsely in the first step and then remove the ambiguities in sentence opinion based on the opinion of document and neighboring sentences. It motivates the design of a coarse-fine opinion mining framework. This framework adopts multi-pass coarse-fine analysis. Similarly, a sentence in a topic-relevant document has higher probability to be relevant and vice versa. Therefore, the coarse-fine analysis mechanism is also applicable to sentence-relevance estimation. The observation on NTCIR-6 corpus and NTCIR-7 training corpus show strong correlation between opinionated sentences and topic-relevant sentences in topic-relevant documents. In NTCIR-6 corpus, 93.1% opinionated sentences in the topic-relevant documents are relevant to the topic while the 73.2% of all of the sentences are topic-relevant. It means that the opinionated sentences in the topic-relevant documents have higher probability to be topic-relevant. Similar correlations are observed in NTCIR-7 training corpus. This motivates the consideration of mutual reinforcement of sentence relevance determination and opinionated sentence classification.

Based on these observation and analysis, a mutual-reinforced and coarse-fine opinion mining framework is designed. The framework is described below.

Input: Document D consists of sentences S_0, S_1,S_i …S_n

Step 1. Use the base classifier for opinion analysis, C_{op_base}, to analyze the opinion of each sentence in D. The output is the polarity value of each sentence, $Pol(S_i)$ with the confidence c_{op}.
Step 2. Use the base classifier for sentence relevance estimation, C_{rel_base}, to estimate the relevance of each sentence in D. The output is the relevance value of each sentence, $Rel(S_i)$ with the confidence c_{rel}.
Step 3. Estimate the polarity of D.

$$Pol(D) = \frac{1}{n} \cdot \sum_{i=1}^{n} Plo(S_i)$$

Step 4. Estimate the topic-relevance of D.

$$\mathrm{Re}l(D) = \frac{1}{n} \cdot \sum_{i=1}^{n} \mathrm{Re}l(S_i)$$

Step 5. Use the complex classifier, C_{op_com}, to estimate the opinion of each sentence, $Pol(S_i)^*$. C_{op_com} incorporates inner-sentence features, sentence-level features and document opinion.
Step 6. Use the complex classifier, C_{rel_com}, to estimate the relevance of each sentence, $Rel(S_i)^*$. C_{rel_com} incorporates inner-sentence features, sentence-level features and document relevance.
Step 7. Adjust the $Pol(S_i)^*$ to $Pol(S_i)^{**}$ according to $Rel(S_i)^*$ and c_{op}. The value of $Pol(S_i)^*$ is increased with a larger $Rel(S_i)^*$, otherwise decreased. The confidence c_{rel} is considered in the adjustment.
Step 8. Adjust the $Rel(S_i)^*$ to $Rel(S_i)^{**}$ according to $Pol(S_i)^*$. The value of $Rel(S_i)^*$ is increased with a larger $Pol(S_i)^*$, otherwise decreased. The confidence c_{op} is considered in the adjustment.
Step 9. Update the document polarity and document relevance using $Pol(S_i)^{**}$ and $Rel(S_i)^{**}$. The confidence values of c_{op} and c_{rel} are increased.
Step 10. If the difference of document polarity and the difference of document relevance after the update lower than a threshold, terminate. Otherwise, go to Step 5.

4. System Implementation

4.1. Preprocessing

Word segmentation and Part-of-Speech tagging are indispensable steps in Chinese sentence analysis. The word segmentation and POS tagging system proposed in [19] are adopted. This system is based Unicode. It is trained using the Peking University People's Daily corpus and Sinica corpus, respectively. Thus, it can process both Traditional Chinese and Simplified Chinese text in one system. Furthermore, the named entity recognizers in [20] are adopted. The recognized name entities are candidates of opinion holders and opinion targets.

The sentiment lexicon is built based on following resources: (a) *The Lexicon of Chinese Positive Words* [21], which consists of 5,054 positive words and the *Lexicon of Chinese Negative Words* [22], which consist of 3,493 negative words; (b) The opinion word lexicon provided by National Taiwan University (NTU) which consists of 2,812 positive words and 8,276 negative words [3]; (c) Sentiment word lexicon and comment word lexicon from Hownet. It contains s 836 positive sentiment words, 3,730 positive comments, 1,254 negative sentiment words and 3,116 negative comment words. These lexicons are encoded in Unicode. The different grapheme corresponding to Traditional Chinese and Simplified Chinese are both considered so that the sentiment lexicons from different sources are applicable to process both Traditional Chinese and Simplified Chinese text. The lexicon is manually verified. Totally, 14,201 positive words, 17,372 negative words and 478 neutral words are obtained. In which, 789 words has more than one polarity. Furthermore, 1,398 strong positive words and 1,983 strong negative words are marked in the lexicon.

4.2. The Base Classifier for Opinion Analysis

The observation on NTCIR-6 corpus shows that our sentiment lexicon achieves 97.3% recall for the opinionated sentences. While further considering the lexicon of opinion operators and opinion indicators [23], the recall increases to 98.5%. Thus, the word-level features are adopted in the base classifier. To increase the classification accuracy, the collocation level features are further incorporated. The employed features are listed below. More description of the features are given in [24].

Table 1. Features adopted in base classifiers for opinion mining

Punctuation level features
The presence of direct quote punctuation “「”and “」”
Word-level and entity-level features
The presence of known opinion operators
The percentage of known opinion word in sentence
The percentage of known strong opinion word in sentence
The presence of a named entity
The presence of pronoun
The presence of known opinion indicators
The presence of known degree adverbs
Collocation-level features
The presence of collocations between named entities and opinion operators

The presence of collocations between pronouns and opinion operators
The presence of collocations between nouns or named entities and opinion words
The presence of collocations between nouns or named entities and strong opinion words
The presence of collocations between pronouns and opinion words
The presence of collocations between pronouns and strong opinion words
The presence of collocations between degree adverbs and opinion words
The presence of collocations between degree adverbs and strong opinion words
The presence of collocations between degree adverbs and opinion operators

The features are linear combined to generate the sentence polarity, $Pol(S_i)$.

4.3. The Base Classifier for Sentence Relevance Estimation

Given a sentence in document *D* of topic *I*. The following features are designed or selected in the base classifier for sentence relevance estimation.

Table 2. Features adopted in base classifiers for sentence relevance estimation

The percentage of named entity in the sentence
The percentage of pronoun in the sentence
The presence of the nouns in the title of the document
The presence of the named entity in the title of the document
The presence of the named entity in the query
The presence of the nouns in the query
The presence of known topic words
The position of the sentence in the document and paragraph. Suppose a document has *p* paragraphs. In the *k-th* paragraph, p_k, has *n* sentences, the position feature of the *i-th* sentence in p_k is estimated by, $$(1-\frac{k-p}{p})+0.5\cdot(1-\frac{i-n}{n})$$
The feature based on the centroid of the document. Suppose there are *N* documents related to topic *i*; and a word *t* appears in one of the document *d* $tf(t,d)$ times and *t* appears in n_t documents of topic *i*. Thus, we define the weight of *t* in the document *d*, labeled as *TF-IDF*(*t*), as $$TF-IDF(t)=tf(t,d)\cdot\frac{N}{n_t}$$ The value of TF-IDF weights the centroid of a word in a document. The centroid of a sentence S_j is then estimated by summing the centroid of each content word in S_j.

The features are linear combined to generate the sentence relevance, $Rel(S_i)$.

4.3. The Complex Classifier for Opinion Analysis

Use the base classifier to analyze the opinion of sentences and document, the coarse analysis results are obtained. Now, we incorporate the document-level and sentence-level features in the complex classifier.

For the *i-th* sentence in the document, labeled as s_i, we assume its polarity, labeled as $Pol(s_i)$, is positive, (its values including positive, neutral, negative and non-opinionated) and the polarity of its previous sentences s_{i-1}, labeled as $Pol(s_{i-1})$, is positive. The conditional probability,

$$P(Pol(s_i)=positive \mid Pol(s_{i-1})=positive)$$
$$=\frac{P(Pol(s_i)=positive \cap Pol(s_{i-1})=positive)}{P(Pol(s_i)=positive)}$$

can be estimated. The conditional probabilities of other polarity co-occurrence combinations between s_i and s_{i-1} are calculated in the same way. Similarly, the conditional probabilities corresponding to the co-occurrences with distance of two sentences are estimated. These conditional probabilities are used as features. Besides the features adopted in the $C_{op\text{-}base}$, the additional features adopted in $C_{op\text{-}com}$ are listed in Table 3.

Table 3. Additional features adopted in complex classifiers for opinion mining

Sentence level features
$P(Pol(s_i)=positive \mid Pol(s_{i-1}))$, values [0,1]
$P(Pol(s_i)=neutral \mid Pol(s_{i-1}))$, values [0,1]
$P(Pol(s_i)=negative \mid Pol(s_{i-1}))$, values [0,1]
$P(Pol(s_i)=non\text{-}opinionated \mid Pol(s_{i-1}))$, values [0,1]
$P(Pol(s_i)=positive \mid Pol(s_{i-2}))$, values [0,1]
$P(Pol(s_i)=neutral \mid Pol(s_{i-2}))$, values [0,1]
$P(Pol(s_i)=negative \mid Pol(s_{i-2}))$, values [0,1]
$P(Pol(s_i)=non\text{-}opinionated \mid Pol(s_{i-2}))$, values [0,1]
$P(Pol(s_i)=positive \mid Pol(s_{i+1}))$, values [0,1]
$P(Pol(s_i)=neutral \mid Pol(s_{i+1}))$, values [0,1]
$P(Pol(s_i)=negative \mid Pol(s_{i+1}))$, values [0,1]
$P(Pol(s_i)=non\text{-}opinionated \mid Pol(s_{i+1}))$, values [0,1]
$P(Pol(s_i)=positive \mid Pol(s_{i+2}))$, values [0,1]
$P(Pol(s_i)=neutral \mid Pol(s_{i+2}))$, values [0,1]
$P(Pol(s_i)=negative \mid Pol(s_{i+2}))$, values [0,1]
$P(Pol(s_i)=non\text{-}opinionated \mid Pol(s_{i+2}))$, values [0,1]
Document level features
Pol(D)

A Support Vector Machine based classifier, which incorporates the features in Table 1 and Table 3, is trained through semi-supervised learning on NTCIR-6 corpus, NTCIR-7 training corpus and more webpage relevant to the documents. The training algorithm is described in [24]. The trained classifier analyzes each input sentence and determines its polarity as the output. Here, the SVM with linear kernel is adopted to perform opinionated sentence identification and polarity determination.

4.4. The Complex Classifier for Sentence Relevance Estimation

Similar to opinion analysis, the document level and sentence level relevance outputted by the base classifier are incorporated in the complex classifier for sentence relevance estimation. The additional document level and sentence level features are listed in Table 4.

A classifier based on linear combination incorporates the listed features in Table 2 and Table 4 to refine the sentence relevance.

Table 4. Additional features adopted in complex classifiers for sentence relevance estimation

Sentence level features
$P(Rel(s_i)=Y \mid Rel(s_{i-1}))$, values [0,1]
$P(Rel(s_i)=N \mid Rel(s_{i-1}))$, values [0,1]
$P(Rel(s_i)=Y \mid Rel(s_{i-2}))$, values [0,1]
$P(Rel(s_i)=N \mid Rel(s_{i-2}))$, values [0,1]
$P(Rel(s_i)=Y \mid Rel(s_{i+1}))$, values [0,1]
$P(Rel(s_i)=N \mid Rel(s_{i+1}))$, values [0,1]
$P(Rel(s_i)=Y \mid Rel(s_{i+2}))$, values [0,1]
$P(Rel(s_i)=N \mid Rel(s_{i+2}))$, values [0,1]
Document level features
Rel(D)

4.5. The Mutual Reinforcement and multi circles

Suppose the polarity value and relevance value of S_i are $Pol(S_i)^*$ and $Rel(S_i)^*$, respectively, the polarity of S_i is adjusted by considering the mutual reinforcement between analysis of sentence opinion and sentence opinion.

$$Pol(S_i)^{**} = MIN\{Pol(S_i)^* \cdot (1 + w_{mu} \cdot \mathrm{Re}\,l(S_i)^*), 1\}$$

where, w_{mu} is mutual reinforcement weight. It is experimentally set to 0.2. The $Rel(S_i)^*$ is adjusted following the similar way.

According to the description in Chapter, the framework is designed as a multi-pass circle. After each circle the confidence weight of w_{op} and w_{rel} are increased. The analysis circles terminate when the output results are converge.

4.6. The Recognition of Opinion Holder and Opinion Target

To recognize the opinion holders, simple co-reference normalization is firstly applied to text in order to recover the bypassed opinion holders in the continuous sentences.

The following heuristics are used to recognize the core of opinion holders:

1. It must be a recognized entity or pronoun.
2. It must collocate and strongly associated with certain identified opinion operators.
3. It always occurs in the beginning of a sentence or near the beginning or end of a quotation.
4. It co-occurred with opinion operators with certain pattern.
5. It frequently co-occurred with the topic words in the query
6. It frequently co-occurred with the entities in the query.

Some heuristics rules and patterns are applied to expand the opinion holder from its core. These manually complied rules and patterns are relevant to punctuations, conjunctions, suffix, prefix and opinion operator. Furthermore, the position of the opinion holder candidate in the sentence and the respective position to the opinion operator candidate are considered.

The opinion targets are not always persons or name entities, they may be nouns, phrases or clauses. For the opinion targets of persons or name entities, the recognition strategy is similar to opinion holder recognition. The corresponding heuristic rules and patterns for opinion target are manually prepared. As for the clause opinion targets, its recognition is highly dependent on the recognition of the opinion operator. The opinion operator always indicates the boundary of clause opinion target. Totally, we manually prepared 25 rules and patterns for opinion targets of persons and 41 patterns for recognizing clause opinion targets. The idea of semantic role labeling is also partially adopted in this subtask.

5. Evaluation

5.1. Datasets

The NTCIR-7 MOAT test corpus is a multilingual comparable corpus across the languages with shared topics. WIA-Opinmine participate the evaluation at Traditional Chinese (*TC*) side and Simplified Chinese (*SC*) side, respectively. The Traditional Chinese data contains data from 1998 to 2001 from the China Times, United Evening News and some other newspapers. The Simplified Chinese data contains documents from Xinhua News and Lianhe Zaobao from 1998 to 2001. TC testing corpus contains a total of 187 documents, 4,665 sentences and 4,668 opinion sub-sentences for 14 topics (Topic 3-16). Corresponding to the same topics, SC testing corpus contains 252 documents and 4,877 sentences. For each side of data, three annotators annotate the opinionated sentences individually. Their outputs are compared to generate the Gold Standards.

5.2. Evaluation Criteria

Five subtasks, including sentence-topic relevance determination, opinionated sentence determination, polarity diction, opinion holder and target identification are evaluated. Among them, sentence-topic relevance and opinion sentence determination adopted the same three metric, i.e. Precision (P), Recall (R) and F.

$$P = \frac{\#system\_current}{\#system\_proposed}$$

$$R = \frac{\#system\_current}{\#gold\_answer}$$

$$F = \frac{2 \times P \times R}{P + R}$$

For the polarity determination, two set of metric are adopted. The first one is Set Precision (S_P) which is defined as,

$$S\_P = \frac{\#system\_corrent(polar = POS, NEU, NEG)}{\#system\_correct(opn = Y)}$$

The second one is recall-based metric. The recall-based precision (R_P) is defined as,

$$R\_P = \frac{\#system\_corrent(polar = POS, NEU, NEG)}{\#system\_proposed(opn = Y)}$$

The recall-based Recall (labeled as R_R) and recall-base F (labeled as R_F) are computed as,

$$R\_R = \frac{\#system\_corrent(polar = POS, NEU, NEG)}{\#gold(opn = Y)}$$

$$R\_F = \frac{2 \times R\_P \times R\_R}{R\_P + R\_R}$$

The evaluation on recognition of opinion holder and opinion target adopts the metric similar to polarity determination.

Considering the inconsistency between three annotators, both strict evaluations and lenient evaluations are conducted. For the strict evaluation, only the annotation outputs agreed by all three annotators are selected to generate the gold standard. The sentences without agreement between all three annotators are not included for evaluation. As for the lenient evaluation, the annotation output agreed by any two of three annotators are included in the gold standard. Corresponding to the gold standard generated by strict and lenient restrictions, the performances of WIA-Opinmine are evaluated, respectively.

5.3. Performance of WIA-Opinmine

Firstly, the sentence-topic relevance determination is evaluated. The achieved precision, recall and F under strict evaluation and lenient evaluation on both TC and SC sides are give in Table 5, respectively.

Table 5. Evaluation of sentence-topic relevance determination

		TC	SC
Strict	P	0.994	0.997
	R	0.530	0.524
	F	0.692	0.687
	Proposed Y	1368	2274
Lenient	P	0.978	0.994
	R	0.406	0.503
	F	0.573	0.668
	Proposed Y	1601	2348

Compared with other participates, the precision of WIA-Opinmine is high but the recall is low. It indicates our current framework should be further improved on recall performance.

Secondly, the opinionated sentence determination is evaluated. The reported performances are given in Table 6.

Table 6. Evaluation of opinionated sentence determination

		TC	SC
Strict	P	0.852	0.609
	R	0.600	0.897
	F	0.704	0.726
	Proposed Y	885	1320
Lenient	P	0.730	0.586
	R	0.521	0.821
	F	0.608	0.683
	Proposed Y	1558	2617

WIA-Opinmine achieves the top-1 or top-2 precisions in both TC side and SC side under lenient evaluation and strict evaluation. Meanwhile, WIA-Opinmine achieves the best F performance on SC side, which shows the high accuracy of our proposed framework.

Thirdly, the performance on polarity determination of opinionated sentences is evaluated. The achieved performances are given in Table 7. Two sets of metrics are adopted here. The first one is Set Precision (labeled as S_P). The second one includes the recalled precision, recall and F (labeled as R_P, R_R and R_F, respectively).

Table 7. Evaluation of polarity determination of opinionated sentences

		TC	SC
Strict	S_P	0.700	0.533
	R_P	0.596	0.325
	R_R	0.420	0.478
	R_F	0.493	0.387
	Evaluated	754	1320
Lenient	S_P	0.699	0.742
	R_P	0.506	0.435
	R_R	0.361	0.609
	R_F	0.421	0.507
	Evaluated	1137	2617

WIA-Opinmine achieves both the best precision and the best F performance on TC and SC side, respectively. Compared with other systems, our system shows the advantage on polarity determination.

Finally, the performance of recognition of opinion holder and opinion target are evaluated. Both lenient evaluation and recall-based lenient evaluation are conducted. The achieved performances are given in Table 8 and Table 9, respectively.

Table 8. Evaluation of Opinion Holder Recognition

		TC	SC
Lenient	P	0.825	0.450
	R	0.825	0.450
	F	0.825	0.450
Lenient Recall-based	P	0.299	0.264
	R	0.430	0.369
	F	0.353	0.308

Table 9. Evaluation of Opinion Target Recognition

		TC	SC
Lenient	P	0.518	0.823
	R	0.518	0.823
	F	0.518	0.823
Lenient Recall-based	P	0.107	0.198
	R	0.479	0.495
	F	0.175	0.283

Four teams provided both opinion holder and opinion target recognition results on TC side, respectively. On the SC side, three teams provided opinion holder recognition results and two teams provided opinion target recognition results. WIA-Opinmine achieves the best precision and the best F on both TC side and SC side. It is shown the effectiveness of our proposed entity-based analysis and holder/target expansion based on heuristic rules.

5.4. Discussions

Comparing with other teams, WIA-Opinmine always achieves better precisions but the recall is not satisfactory. It shows that the proposed coarse-fine opinion mining framework is good at high precision. It should be enhance the recall performance. In the three of five sub-tasks, WIA-Opinmine achieves better F performance including polarity determination, opinion holder recognition and opinion target recognition. On the contrary, the F performance on sentence relevance determination and opinionated sentence determination are not satisfactory. This result partially attributes the unsatisfactory performance of the classifiers for sentence relevance determination since it is not well studied. Meanwhile, the mutual reinforcement between sentence relevance and opinionated sentence influences the opinionated sentence classification since the unsatisfactory performance of sentence relevance determination. It means that the recognition errors on one side have the risk to affect the other side of mutual reinforcement. Fortunately, the achieved best performance on polarity determination supports such consideration of mutual reinforcement.

6. Conclusions

In this paper, we present WIA-Opinmine system in NTCIR-7 MOAT task. The system adopts a coast-fine analysis strategy in opinion mining. The multi-pass coast-fine analysis utilizes the document opinion and neighboring sentence opinions to incrementally refining the sentence opinion analysis. Meanwhile, the mutual reinforcement of sentence relevance and sentence opinion analysis use the analysis results on one side to help the analysis on the other side. The evaluations on Traditional Chinese side and Simplified Chinese side in NTCIR-7 MOAT show the effectiveness of the proposed coast-fine opinion analysis framework. The future researches will focus on the recall enhancement.

References

[1] M. Q. Hu and B. Liu, Opinion Extraction and Summarization on the Web, In *NCAI06*, Boston, 2006

[2] H. Yu and V. Hatzivassiloglou, Towards Answering Opinion Question: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences, In *EMNLP03*

[3] L.W. Ku, T.H. Wu, L.Y. Lee, H.H. Chen, Construction of an Evaluation Corpus for Opinion Extraction, In *NTCIR-5*, pp.513-520, Japan, 2005

[4] P. Chaovalit, L. Zhou, Movie Review Mining: a Comparison between Supervised and Unsupervised Classification Approaches, In *HICSS05*, 2005

[5] Y. Seki, D.K.Evans, L.W. Ku, H.H. Chen, N. Kando and C.Y. Lin, Overview of Opinion Analysis Pilot Task at NTCIR-6, In *NTCIR-6*, pp.456-463, Japan, 2007

[6] Y. Seki, D.K.Evans, L.W. Ku, Overview of Multilingual Opinion Analysis Task at NTCIR-7, In *NTCIR-7*,2008

[7] V. Hatzivassiloglou and K. McKeown K., Predicting the Semantic Orientation of Adjectives. In *ACL1997*, pp.174-181, Madrid, 1997

[8] P. Turney, M. Littman, Measuring Praise and Criticism: Inference of Semantic Orientation from Association. *ACM Trans. Information Systems*, 4(21), 315-346, 2003

[9] J. Kamps, M. Marx, R.J. Mokken, M. Rijke, Using WordNet to Measure Semantic Orientation of Adjectives. In *LREC04*, 2004

[10] V. Hatzivassiloglou and J. M. Wiebe. Lists of manually and automatically identified gradable, polar, and dynamic adjectives. www.cs.pitt.edu/wiebe/pubs/coling00/coling00adjs.tar.gz

[11] M.Q. Hu and B. Liu, Mining and Summarizing Customer Reviews, In *ACM SIGKDD 2004*, pp.168-177, 2004

[12] W. Zhang, S. Liu and W.Y Meng, Opinion Retrieval from Blogs. In *ACM CIKM07*, pp.831-840, Lisbon, 2007

[13] J. Wiebe, T. Wilson, R. Bruce, Learning Subjective Language. *Computational Linguistics*, 30[3]: 277–308, 2004

[14] R.H. Huang et al. ISCAS in Opinion Analysis Pilot Task: Experiments with Sentimental Dictionary based on Classifier and CRF model. In *NTCIR-6*, Japan, pp.296–300, 2007

[15] B. Pang, L.L. Lee and S. Vaithyanathan, Thumbs up? Sentiment Classification Using Machine Learning Techniques. In *EMNLP02*, pp.79-86, Philadelphia, 2002

[16] B. Pang and L.L. Lee, A Sentiment Education: Sentiment Analysis Using Subjectivity Summarization based on Minimum Cuts. In *ACL04*, pp. 271-278, Spain, 2004

[17] Riloff, J. Wiebe and T. Wilson, Learning Subjective Nouns Using Extraction Pattern Bootstrapping. In *CoNLL03*, pp.25-32, 2003

[18] .M. Kim and E. Hovy, Extracting Opinions Expressed in Online News Media Text with Opinion Holders and Topics," In *COLING-ACL06*, 2006

[19] Q. Lu, S. T. Chan, R. F. Xu, et al. A Unicode based adaptive segmentor. In *2nd SIGHAN Workshop on Chinese Language Processing at ACL 2003*, pp.164-167, Spain, 2003

[20] G. H. Fu and K. K. Luke. Chinese POS disambiguation and unknown word guessing with lexicalized HMMs. *International Journal of Technology and Human Interaction*, Vol.2, No.1, pp.39-50, 2006

[21] J.L. Shi and Y.G. Zhu, *Lexicon of Chinese Positive Words*, SiChuan Dictionary Press, 2005

[22] L. Yang and Y.G. Zhu, *Lexicon of Chinese Negative Words*, Sichuan Dictionary, 2005

[23] R.F. Xu, K.F. Wong and Y.Q. Xia, Opinmine – Opinion Analysis System by CUHK for NTCIR-6 Pilot Task. In *NTCIR-6 Workshop*, Japan, pp.350-357, 2007

[24] R.F, Xu, K.F. Wong, Q. Lu and Y.Q. Xia, Learning Knowledge from Relevant Webpage for Opinion Analysis, In *Proc. IEEE/WIC/ACM WI-IAT*, 2008.