Sentence-Level Opinion Analysis by CopeOpi in NTCIR-7

Lun-Wei Ku, I-Chien Liu, Chia-Ying Lee, †Kuan-hua Chen and Hsin-Hsi Chen
Department of Computer Science and Information Engineering
†Department and Graduate Institute of Library and Information Science
National Taiwan University
Taipei, Taiwan
{lwku, icliu, cylee}@nlg.csie.ntu.edu.tw; khchen@ntu.edu.tw; hhchen@csie.ntu.edu.tw

Abstract

In this paper, we introduce our system, CopeOpi, for analyzing opinionated information in NTCIR-7 MOAT task's document collections. We participated in all tasks except opinion target extraction and submitted three runs for both simplified and traditional Chinese sides. For opinion extraction task, our algorithm was based on the bag-of-character methods proposed in NTCIR-6 and considered morphological structures of Chinese words to extract opinion words more correctly. How distant an opinion word is to the end of the sentence is also considered to adjust its opinion weight. The performance of the opinion extraction, which is the second best of all participants, achieves the f-measure 0.672 under the lenient metric and 0.783 under the strict metric. The performances of polarity detection and the relevance judgment are both ranked the third.

Keywords: *Opinion Extraction, Opinion Polarity Detection, Opinion Holder Extraction, Sentence Relevance Judgment, Sentiment Mining*

1 Introduction

Opinion analysis has drawn a widely attention in these days. In addition to facts, people are concerned about opinions, and practical applications are also proposed in many areas, such as market investigation, public poll, etc. Generally speaking, an opinion is composed of the holder, the expressed attitude, and the target. To analyze opinions, we have to extract them first. Then we have to identify their holders and targets to realize who expressed these opinions and what were they talking about. After that, we need to keep only those sentences which are relevant to our focus. Also to know the attitude, supportive or non-supportive, is also important for deep opinion mining.

Several techniques are indispensable for the opinion analysis task. To extract opinions, machine learning methods and lexical pattern extraction methods were both adopted by researchers[6]. Dictionaries and other resources were constructed [2]. To judge the relevant sentences, techniques in information retrieval were explored. To extract holders, features and rules are discussed [4][5]. The concept of target extraction has been combined with relevancy judgment, and it is a new task in NTCIR-7 MOAT. Multilingual opinion analysis, the goal of MOAT, is also getting important because of the global information exchange, and it becomes a new research direction [7].

We participated in opinion extraction, polarity detection, sentence relevance judgment, and opinion holder extraction tasks at both simplified and traditional Chinese sides in NTCIR-7 MOAT cluster. In this paper, our system CopeOpi is proposed and our methods are described. Discussions of the performance are also included according to the experimental results to improve the system in the future.

2 An Chinese Opinion Extraction System: NTU CopeOpi

The **C**hinese **op**inion **e**xtraction system for **op**inionated **i**nformation (CopeOpi) is a web-based system developed from news documents. It can extract sentiment words, sentences and documents. Moreover, it ranks the importance of documents by the informative degree of their possessive sentences.

The components for extracting opinion words and sentences, and then decide their opinion polarities in this system were used as the basic components in MOAT tasks. Moreover, we added some new features and modified some formulae to these basic components to improve the system. These features and new formulae will be discussed in the following sections.

3 Opinion Extraction

Three factors were considered when extracting opinion sentences and determining their polarities: the sentiment words, the opinion operators, and the negation operators. Only when one sentence was judged as opinionated, its opinion holder(s) was reported by the system.

In NTCIR-6, we postulated that the opinion of the whole is a function of the opinions of the parts. That is, the opinion degree of a sentence, which decides if this sentence is opinionated, and a sentence's polarity, is a function of sentiment words, negation words, and opinion operators. Negation words (41) and opinion operators (151) were collected in word lists. Then we calculated the sentiment score of each Chinese character by their observed probabilities [2]. To recognize sentiment words, the opinion score of each word were calculated using the bag-of-character method and morphological-structure method. One positive threshold and one negative threshold were set. If the opinion score of one word was greater than the positive threshold, it was reported as a positive opinion word; if the score was less than the negative threshold, this word was reported negative. The word whose score falls within was reported neutral, and the word whose score equals zero was reported non-opinionated.

3.1 Opinion Score of Words

Sentiment words are employed to compute the tendency of a sentence. A Chinese opinion dictionary NTUSD [2] was adopted. NTUSD consists of 2,812 positive and 8,276 negative opinion words. If a word was found in NTUSD, its polarity was reported as the polarity recorded in NTUSD, and its weight was calculated by CopeOpi. If we couldn't find the word in NTUSD, its polarity and weight were both determined by CopeOpi.

As we mentioned in section 2, we calculated characters' opinion scores by methods adopted in NTCIR-6. Based on them, we calculated the words' opinion scores by three methods in NTCIR-7: bag-of-character method in NTCIR-6 (run 2), morphological method (run 3), morphological method plus position weight (run 1). In this section, we first introduce morphological method. The position weight will be discussed later in section 3.3.

The morphological method combines the opinion scores of a word's composite characters according to its morphological type. In the Chinese language, five morphological types are defined by linguistics [1], and we defined three more types for opinion analysis.

(1) Parallel Type: Two morphemes play coordinate roles in a word. For example, the morphemes “財” (money) and “富” (wealth) are parallel in the word “財富” (money-wealth).

(2) Substantive-Modifier Type: A modified morpheme follows a modifying morpheme. For example, the morpheme “級” (class) is modified by “低” (low) in the word “低級” (low-classed).

(3) Subjective-Predicate Type: One morpheme is an expresser and the other is described. The structure is like a subject-verb sentence condensed in one word. For example, the morpheme “心” (heart) is a subject of the predicate “疼” (hurt) in the word “心疼” (heart-hurt).

(4) Verb-Object Type: The first morpheme is usually a verb which governs the second one, making this word similar to a verb followed by its object. For example, the morpheme “控” (control) serves as the object of the verb “失” (lose) in the word ”失控” (lose-control).

(5) Verb-Complement Type: The first morpheme is usually a verb but sometimes can be an adjective, and the second morpheme explains the first one from different aspects. For example, the morpheme “清” (clearly) expresses the aspect of the action “看” (look) in the word “看清” (look-clearly).

Type (6) and (7) include words which have special characteristics when calculating opinion scores, therefore they are separated from Type (1) to (5). Type (6) defines words whose prefix or suffix means negation and Type (7) confirmation. Words whose prefix or suffix characters are not morphemes, i.e., the remaining three cases, are classified into Type (8).

(6) Negation Type: There is at least one negation character in the word of this type. For example, the prefix “無” (no) and “不” (not) are negation morphemes in the words ”無法” (no-method) and ”不慎” (not-careful).

(7) Confirmation Type: There is at least one confirmation character in the word of this type. For example, the prefix “有” (do; have) is a confirmation morpheme in the words “有賴” (do-depend on).

(8) Others: Those words that do not belong to the above seven types are assigned to this type, such as words whose meanings are not a function of their composite characters, words whose composite characters are not morphemes, and so on. For example, “姪子” (nephew-suffix), “阿媽” (prefix-grand mother) and “薄荷” (peppermint).

After introducing eight morphological types, their corresponding formulae for calculating opinion scores are listed below.

(1) Parallel Type: Since the two composite characters of a word of this type are homogeneous, the opinion score is the average scores of the two characters. For example, the word “疲乏” (tired) is composed of “疲” (tired, negative) and “乏”

(exhausted, negative). Its opinion score is determined by the average scores of "疲" and "乏".

$$S(C_1C_2) = \frac{S(C_1) + S(C_2)}{2} \quad (1)$$

(2) Substantive-Modifier Type: The first morpheme of a word of this type modifies the second one, so that its opinion weight comes from the absolute opinion score of the first character, while the opinion polarity is determined by the occurrence of negative opinion characters. If at least one negative opinion character appears, then the word is negative, else it is positive. For example, the word "痛哭" (bitterly cry) is composed of "痛" (bitterly, negative) and "哭" (cry, negative). This word is negative because there are negative characters in it. Its opinion strength, i.e., the absolute value of the opinion score, is decided by the first character "痛" (bitterly), which describes the degree of crying. Words such as "浩劫" (disaster, negative) composed of "浩" (great, positive) and "劫" (calamity, negative) and "高強" (excellent, positive) composed of "高" (superior, positive) and "強" (force, positive) are of this type, but different combinations.

$$\begin{array}{l} \text{if } (S(C_1) \neq 0 \text{ and } S(C_2) \neq 0) \\ \quad \text{if } (S(C_1) > 0 \text{ and } S(C_2) > 0) \\ \quad\quad S(C_1C_2) = S(C_1) \\ \quad \text{else } S(C_1C_2) = -1 \times |S(C_1)| \\ \text{else } S(C_1C_2) = S(C_1) + S(C_2) \end{array} \quad (2)$$

(3) Subjective-Predicate Type: The first morpheme of a word of this type is a subject and the second morpheme is the action it performs, so that the action decides the opinion score of the word. If the action is not opinionated or it is neutral, the subject determines the opinion score of this word. For example, the word "山崩" (mudslide, negative) is composed of "山" (mountain, non-opinionated) and "崩" (collapse, negative). Its opinion score depends only on the second character "崩" (collapse) since the first character is a subject and usually bears no opinions.

$$\text{if } (S(C_2) \neq 0) \text{ then } S(C_1C_2) = S(C_2) \text{ else } S(C_1C_2) = S(C_1) \quad (3)$$

(4) Verb-Object Type: The first morpheme of a word of this type acts upon the second morpheme. The effect depends not only on the action but also on the target. The weight is determined by the action, but the opinion polarity is the multiplication of the signs of the two morphemes. For example, the word "避暑" (to go away for the summer, positive) is composed of "避" (hide, negative) and "暑" (hot summer, negative). Its opinion strength depends on the strength of "避" (hide) and the opinion polarity is the multiplication of two negatives, so that positive is derived. The word "抗病" (disease-resistant, positive) composed of the character "抗" (resist, negative) and "病" (disease, negative) also belongs to this type.

$$\begin{array}{l} \text{if } (S(C_1) \neq 0 \text{ and } S(C_2) \neq 0) \\ \quad \text{then } S(C_1C_2) = |S(C_1)| \times SIGN(S(C_1)) \times SIGN(S(C_2)) \\ \quad \text{else } S(C_1C_2) = S(C_1) + S(C_2) \end{array} \quad (4)$$

(5) Verb-Complement Type: The scoring function for a word of this type is defined the same as that of a Subjective-Predicate type, i.e., Formula (3). The complement morpheme is the deciding factor of the opinion score. For example, the word "提高" (raise, positive) is composed of "提" (carry or lift, non-opinionated) and "高" (high, positive). The complement morpheme "高" (verb) describes the resulting state of the verb morpheme "提" (raise), therefore both opinion strength and polarity depend on the complement morpheme "高" (high).

(6) Negation Type: A negative character specified in a predefined set *NC* has a negation effect on the opinion score of the other character. For example, the word "不悅" (unhappy, negative) is composed of the negative morpheme "不" (not) and the modified morpheme "悅" (pleased, positive). The strength depends on the modified morpheme "悅" (pleased) while the polarity of the word is the negation of the polarity of the modified morpheme.

$$\begin{array}{l} \text{if } (C_1 \in NC) \text{ then } S(C_1C_2) = (-1) \times S(C_2) \\ \text{else } S(C_1C_2) = (-1) \times S(C_1) \end{array} \quad (5)$$

(7) Confirmation Type: A positive character specified in a predefined set *PC* ensures that the opinion score of a word comes from the other character. In other words, it takes no effect in deciding the opinion score of a word. For example, the word "有利" (have profits) is composed of the positive morpheme "有" (have) and the modified morpheme "利" (profits, positive). The opinion score of this word is determined by the modified morpheme "利" (profits).

$$\text{if } (C_1 \in PC) \text{ then } S(C_1C_2) = S(C_2) \text{ else } S(C_1C_2) = S(C_1) \quad (6)$$

(8) Others: Since words of this type contain no clear cues for their morphological structures, we postulate that both characters have the same contribution, and adopt Formula (1).

The magnitude of the opinion score of an unknown word is also the indication of whether it should be counted. In our system, if a word does not appear in the dictionary, that is, it is unknown, only the word whose opinion score is above 0.3 or below –0.3 is taken into consideration, i.e. treated as a sentiment word. Note that the opinion score of one word could vary from –1.0 to 1.0.

3.2 Possible Sentiment Words

The scoring functions were not applied to all words. Parts of speech of words are considered in extracting possible sentiment words. In CopeOpi, only words with part of speech A (adjective), V (verb), Na (proper noun), D (adverb) and Cbb (conjunction) are selected for further calculations of opinion scores. From observations, Chinese words are mostly multi-character words, and one character itself usually cannot express a complete concept. Hence single-character words were not considered opinionated.

3.3 Negation Operator and Position Weight

Negation operators are words such as "不" (no), "沒有" (not), "從不" (never), "也不" (neither), "不可能" (impossible), etc.. These words reverse the meanings of sentences. Moreover, if they modify sentiment words, the opinion polarities of these sentiment words will be reversed, too.

As mentioned, 41 negation operators are collected. For each sentence, after assuring that words are opinionated by the formula in section 3.1, each negation operator will negate the opinion polarity of the closest sentiment word, that is, change the opinion score of that word from S to –S. The effect of a negation operator will not cross commas, periods, question marks, semicolons, and exclamation marks. Sentence segments separated by these punctuation marks are referred to as "sentence fragments". Negation operators themselves can also express negative attitudes. Therefore, if there are no sentiment words in one sentence fragment, the scores of the negation operators within are counted.

Position weight adjusts the opinion score of one word according to its position in one sentence. The idea of this weight is inspired by the document structure. In one document, if a sentence is located in the end of the document, it may summarize the document and thus it is important. We postulated that if one word is close to the end of a sentence, it may be of more importance, too. Therefore we re-weighted the opinion score of words by Formula (7).

$$new\_score = score + \frac{position}{length} * score \tag{7}$$

The setting of three runs for opinion extraction and polarity detection are as follows: (Morphological information M, Observed probability O, Distance information D)

Run 1: M + O + D
Run 2: O
Run 3: M + O

3.4 Opinion Holder Extraction

The opinion holder of an opinionated sentence is usually composed of many words. In the training phase, an opinion holder was first segmented into several words if possible to fit its format in the pre-processed testing data. For example, "美國總統柯林頓" (The U.S. president Clinton) was divided into "美國" (U.S.), "總統" (president), and "柯林頓" (Clinton). We treated this problem as a binary classification problem of determining weather a word in an opinionated sentence is a part of an opinion holder, and trained a classification model with training instances. In the testing phase, we selected the most confident word which is identified as a part of an opinion holder, and combined it with the word next to it by some rules to generate the final opinion holder. Then an opinion holder was reported. A decision tree algorithm "CHAID" provided in the tool "RapidMiner" are adopted to the classification task [3]. Table 1 lists the features we used for training our binary classifier.

Lexical features	
POS	the part of speech of the word
Is_Location	whether the word is a location name
Is_Organize	whether the word is an organization name
Is_Person	whether the word is a person name
Is_Pronoun	whether the word is a pronoun
Is_Noun	whether the word is a noun
Nearest_Verb	the nearest verb in this sentence
Keyword related features	
Has_Operator	whether any opinion operators are in this sentence
Has_PosKW	whether any positive keywords are in this sentence
Has_NegKW	whether any negative keywords are in this sentence
Has_NeuKW	whether any neutral keywords are in this sentence
Nearest_PosKW	the nearest positive keyword in this sentence
Dist_PosKW	the distance from the word to the nearest positive keyword
The other features	
Near_Sen_Start	the word is the first or the second word in this sentence
After_Paren	the word is right next to a parenthesis, e.g., "」", "】"
Before_Colon	the word is left next to the colon, e.g., "：", ":"
Ever_Holder	the word was a part of an opinion holder in previous sentences
Word_Order	the order of the word in this sentence

Table 1. Features for training our binary classifier

We manually collected the opinion operators. Positive, neutral, and negative keywords were collected from NTUSD.

The CHAID model predicted YES to a part of the opinion holder or NO to other words, and a confidence score to this prediction. We first found the word with the highest confidence. Then we considered the part of speech information of its surrounding words to generate the complete opinion holder. This merge process includes three steps:

Step 1: We combineed this word with the coming nouns of part of speech “Na” (common nouns), “Nb”(proper nouns), or “Nc” (location nouns).

Step 2: If there were words of part of speech Ca (coordinate conjunctions) or “的” (nominal markers) following the current opinion holder, we combined the current holder and the “Ca” or “的”. We repeated from Step 1 until there was nothing to combine.

Step 3: If the holder was a unigram, discarded it.

The classifier may predict “NO” to all words in one opinionated sentence. If there was no opinion holder after the above three steps, “POST_AUTHOR”, the author of the article, will be reported as the default opinion holder by our system.

Three runs were submitted for opinion holder extraction task, and they were different only in the training instances. Three training sets of different sizes were adopted, including NTCIR-7 training set (Run 1), NTCIR-7 training set plus NTCIR-6 positive cases (Run 2), and 17,000 instances in NTCIR-7 training set (Run 3).

3.5 Relevance Sentence Judgment

A rank approach was adopted in the sentence relevance judgment task. The basic idea of this method is to generate a ranked list of sentences by scoring functions, and then determines a cutoff threshold. Sentences whose scores were lower than the cutoff threshold were viewed as irrelevant.

Our scoring functions considered both similarity scores and distance scores. The similarity score indicates how relevant a sentence is to the topic, while the distance score indicates how close a sentence is to the most relevant sentence in the same document. These two scores are summed to generate the final similarity score of a sentence. Sentences are then ranked by their final scores in a descending order. Since all the input documents are relevant to the topic, we assumed that there is at least one relevant sentence in each document. Under this assumption, the cutoff threshold was set at the largest value which ensures that there was at least one relevant sentence in each document.

We adopted Lucene[1] as our basic IR system. We indexed each sentence as a "document" in Lucene. After indexing, we acquired the weight of each term from Lucene, so that each sentence can be represented as a vector of composite terms by their weights. We also generated a centric vector to represent the topic by the summation of all sentential vectors. Then the similarity score of a sentence to a topic is defined as the inner product of the sentential vector and the centric vector.

We applied two different scoring functions to compute distance scores. The first one DF_1 computes the distance between the measured sentence and the most relevant sentence, which has the max similarity score to the topic, in the same document. The distance score is simply set to the inverse of this distance. In the second function DF_2, the distance score of one sentence is contributed by all other sentences. It is defined as the summation of the similarity score divided by the distance from all other sentences in one document in formula (8).

In formula (8), pos(d, s_i) is the position of sentence s_i in document d, and distance dis(s_i, s_j) is the absolute value of pos(s_i) minus pos(s_j).

$$s_{\max} = \max_{s_i}(similarity\_score(s_i))$$

$$DF_1(S_i) = \frac{1}{dis(s_i, s_{\max})} \qquad (8)$$

$$DF_2(S_i) = \sum_{s_j \in d} \frac{similarity\_score(s_j)}{dis(s_i, s_j)}$$

Since the results are not satisfied, we did some modifications on the centric vector. Because weights of many terms in the centric vector are very small, we treated them as insignificant terms or noises. Then a “halved” centric vector is generated by setting those weights less than the median of all weights to zeros. As a result, half of the values in the halved centric vector are zeros and the others are original values. This halved centric vector of each topic is used to compute the similarity score discussed above. The settings of three runs for the relevance judgment task are as follows, and the procedure of the relevance judgment is shown in Figure 1.

Run 1: centric vector + DF1

Run 2: centric vector + DF2

Run 3: halved centric vector + DF1

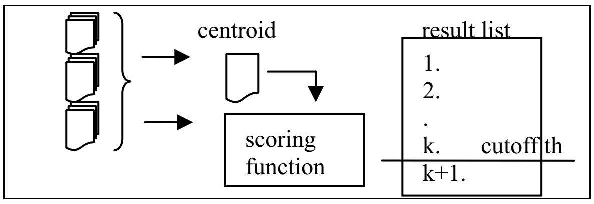

Figure 1. Relevance Judgment

3.6 Simplified Chinese Tasks

For simplified Chinese task, we first translated the corpus to traditional Chinese. Then those techniques described in the previous sections are applied to the simplified corpus. Different settings were set only for the opinion extraction task. In this task, we gave an extra restriction that only sentences containing opinion operators could possibly be opinionated.

4 Experiments and Discussions

The experiment results are shown in Table 2, 3, 4, 5 and 6. Table 2, 3, and 6 show the results of traditional Chinese tasks, and Table 4, 5 and 6 show the results of simplified Chinese tasks. For opinion sentence extraction, our system is ranked the second

[1] http://lucene.apache.org/

among participants. The recall of our system is obviously higher than the precision, which may indicate that our system extracted too many opinion sentences. These results conformed to our observations. Because our system extracts opinion sentences based on opinion words, whenever a word in one sentence is judged opinionated, this sentence also is easily judged opinionated. Especially when there are some weak opinionated words in one sentence but it turns out to be non-opinionated, or several opinionated words with different polarities and they cancelled each other in one sentence. Our system can easily identify opinionated sentences, but has difficulties in weak opinions or neutral sentences.

The results further show that performance of the polarity detection of our system, ranked the third, performs not as well as in opinion extraction. That is, the calculated opinion score may not correctly represent the opinion weight of words. This problem may be the reason of poor performance in detecting neutral and weak opinionated sentences.

The issue of the inter-words modification may be the reason of non-precise opinion weights. We still adopted a bag of words model, and similar to the bag of character model, the modifications are not considered in it. To clarify the modifications among words of different opinion polarities and design of learn the scoring functions according to them may be a feasible direction to improve the performance of the polarity detection task.

For results shown in Table 3 and 5, we found that we extracted too few relevant sentences. In other words, the cutoff threshold is too high. We have mentioned that we set the cutoff threshold to the largest value which guarantees that at least one sentence in each document is relevant. The disadvantage of this method is that there is always a document having only one relevant sentence and this criterion may be too strict. We may need to relax our restriction. This problem gets obvious when there are many relevant sentences [8].

The performance of the opinion holder extraction task is not as good as we expected. Some sorts of holders are not extracted correctly in our system. For example, long compound opinion holders which are composite of several words, foreign opinion holders, and common nouns as opinion holders. The difficulty in extracting long compound opinion holders lies in identifying their boundaries. Foreign names are prone to be segmented incorrectly before processing. When we extracted opinion holders, they were very often only partially correct if they are not properly segmented. A possible way to solve this problem is to add the translation rules of foreign names. Besides, some foreign names were even not translated into Chinese and were tagged as foreign words, for example, TowerGroup. We cannot extract them as holder candidates according to their parts of speech. More context information should be considered in this case. Common nouns, such as “資訊業” (computer industry), are not properly detected by our existing features. Lexical features may not be enough to extract them. Semantic information, such as occupations, titles, nouns which can denote a person, should be included in our feature set to identify common-noun holders.

The performances of simplified Chinese were similar except in the holder extraction task. This may be due to the common usage of terms is different in traditional and simplified Chinese. Our algorithm in opinion holder extraction task included lexical rules. These rules depend on the sentence structure, may not directly be applied to the simplified Chinese corpus.

Group	RunID	L/S	Opinionated			Polarity-T	Polarity-RB		
			P	R	F	Set Precision	P	R	F
CHUK	1	L	0.7305	0.4823	0.5810	**0.6973**	0.5093	0.3363	**0.4051**
CityUHK	1	L	0.6598	0.7820	**0.7158**	**0.5384**	0.3552	0.4210	**0.3853**
CityUHK	2	L	0.7427	0.6045	**0.6665**	**0.5218**	0.3876	0.3155	**0.3478**
CityUHK	3	L	0.6520	0.8054	**0.7206**	**0.5048**	0.3291	0.4066	**0.3637**
iclpku	1	L	0.7006	0.5807	0.6350	0.4824	0.3380	0.2801	0.3063
iclpku	2	L	0.5806	0.6837	0.6280	0.4508	0.2617	0.3082	0.2831
NLCL	1	L	0.5344	0.2478	0.3386				
NLCL	2	L	0.4757	0.6880	0.5625	N/A			
NLCL	3	L	0.4937	0.4696	0.4813				
*NTUCopeOpi	1	L	0.5643	0.8297	**0.6717**	**0.4885**	0.2756	0.4053	**0.3281**
*NTUCopeOpi	2	L	0.5569	0.8203	0.6635	0.4811	0.2679	0.3946	0.3192
*NTUCopeOpi	3	L	0.5569	0.8203	0.6635	0.4826	0.2688	0.3959	0.3202
**TTRD	1	L	0.5103	0.8646	0.6418	0.3732	0.1905	0.3227	0.2395
TTRD	2	L	0.5663	0.6130	0.5887	0.4632	0.2623	0.2840	0.2727
UniNe	1	L	0.5418	0.8574	**0.6640**	0.4295	0.2327	0.3682	0.2852
CHUK	1	S	0.8533	0.5663	0.6808	**0.7063**	0.6027	0.4000	**0.4809**
CityUHK	1	S	0.8373	0.8517	**0.8444**	**0.5497**	0.4602	0.4682	**0.4642**
CityUHK	2	S	0.9005	0.6914	**0.7822**	**0.5309**	0.4780	0.3670	**0.4153**
CityUHK	3	S	0.8186	0.8689	**0.8430**	**0.5259**	0.4305	0.4569	**0.4433**
iclpku	1	S	0.8577	0.6592	0.7454	**0.5091**	0.4366	0.3356	0.3795
iclpku	2	S	0.7431	0.7408	0.7419	0.4904	0.3644	0.3633	0.3638
NLCL	1	S	0.6276	0.2764	0.3838	N/A			

NLCL	2	S	0.5842	0.6989	0.6364				
NLCL	3	S	0.6015	0.4951	0.5431				
*NTUCopeOpi	1	S	0.7082	0.8764	**0.7834**	0.4991	0.3535	0.4375	**0.3910**
*NTUCopeOpi	2	S	0.6984	0.8637	0.7723	0.4848	0.3386	0.4187	0.3744
*NTUCopeOpi	3	S	0.6984	0.8637	0.7723	0.4883	0.3410	0.4217	0.3771
**TTRD	1	S	0.6459	0.8854	0.7469	0.3909	0.2525	0.3461	0.2919
TTRD	2	S	0.7389	0.6360	0.6836	0.4735	0.3499	0.3011	0.3237
UniNe	1	S	0.6921	0.8839	**0.7763**	0.4449	0.3079	0.3933	0.3454

Table 2. Traditional Chinese opinion analysis results

Group	RunID	Relevance (L)			Relevance (S)		
		P	R	F	P	R	F
CHUK	1	0.9787	0.3779	0.5452	0.9949	0.5002	0.6657
iclpku	1	0.9529	0.5231	0.6754	0.9943	0.6375	0.7769
iclpku	2	0.9529	0.5231	0.6754	0.9943	0.6375	**0.7769**
NLCL	1	0.8487	0.1355	0.2338	0.9240	0.1700	0.2872
NLCL	2	0.8578	0.3832	0.5297	0.9290	0.4575	0.6131
NLCL	3	0.8640	0.2670	0.4079	0.9298	0.3217	0.4780
*NTUCopeOpi	1	0.8848	0.5997	**0.7149**	0.9614	0.6699	**0.7896**
*NTUCopeOpi	2	0.9198	0.5484	**0.6871**	0.9804	0.6080	0.7506
*NTUCopeOpi	3	0.9122	0.5446	0.6820	0.9806	0.5962	0.7416
**TTRD	1	0.8992	0.7482	**0.8168**	0.9672	0.7932	**0.8716**
TTRD	2	0.8883	0.8072	**0.8458**	0.9659	0.8447	**0.9012**
UniNe	1	0.8746	0.7855	**0.8276**	0.9614	0.7972	**0.8716**

Table 3. Traditional Chinese relevance extraction results

<table>
<tr><th rowspan="2">Group</th><th rowspan="2">RunID</th><th colspan="2">Lenient</th><th colspan="2">Strict</th></tr>
<tr><th>Opinionated</th><th>Polarity-RB</th><th>Opinionated</th><th>Polarity-RB</th></tr>
<tr><td>BUPT</td><td>1</td><td>0.4807</td><td>N/A</td><td>0.5200</td><td>N/A</td></tr>
<tr><td>ICLPKU</td><td>1</td><td>0.6003</td><td>0.2705</td><td>0.5801</td><td>0.1645</td></tr>
<tr><td>ICLPKU</td><td>2</td><td>0.5745</td><td>0.2599</td><td>0.5373</td><td>0.1508</td></tr>
<tr><td>NEUNLP</td><td>1</td><td>0.5676</td><td rowspan="8">N/A</td><td>0.5469</td><td rowspan="8">N/A</td></tr>
<tr><td>NLCL</td><td>1</td><td>0.4197</td><td>0.3937</td></tr>
<tr><td>NLCL</td><td>2</td><td>0.4178</td><td>0.4144</td></tr>
<tr><td>NLCL</td><td>3</td><td>0.5336</td><td>0.4827</td></tr>
<tr><td>NLPR</td><td>1</td><td>0.6650</td><td>0.7240</td></tr>
<tr><td>NLPR</td><td>2</td><td>0.5311</td><td>0.5798</td></tr>
<tr><td>NLPR</td><td>3</td><td>0.5071</td><td>0.5060</td></tr>
<tr><td>NLPR</td><td>4</td><td>0.5703</td><td>0.6207</td></tr>
<tr><td>NTU</td><td>1</td><td>0.6013</td><td>0.2980</td><td>0.6863</td><td>0.2318</td></tr>
<tr><td>NTU</td><td>2</td><td>0.6011</td><td>0.3053</td><td>0.6871</td><td>0.2481</td></tr>
<tr><td>NTU</td><td>3</td><td>0.6011</td><td>0.3101</td><td>0.6871</td><td>0.2563</td></tr>
<tr><td>TTRD</td><td>1</td><td>0.5772</td><td>0.2510</td><td>0.5124</td><td>0.1476</td></tr>
<tr><td>TTRD</td><td>2</td><td>0.5607</td><td>0.2774</td><td>0.5193</td><td>0.2038</td></tr>
<tr><td>*ISCAS</td><td>1</td><td>0.5723</td><td>N/A</td><td>0.5597</td><td>N/A</td></tr>
</table>

Table 4. Simplified Chinese opinion analysis results

Group	RunID	Relevance (L)			Relevance (S)		
		P	R	F	P	R	F
ICLPKU	1	0.9775	0.6559	0.7850	0.9845	0.6743	0.8004
ICLPKU	2	0.9775	0.6559	0.7850	0.9845	0.6743	0.8004
NLCL	1	0.9630	0.3258	0.4869	0.9736	0.3326	0.4959
NLCL	2	0.9752	0.2799	0.4349	0.9848	0.2846	0.4415
NLCL	3	0.9714	0.5850	0.7302	0.9827	0.5897	0.7371
NLPR	1	0.9510	1.0000	0.9749	0.9633	1.0000	0.9813
NLPR	2	0.9510	1.0000	0.9749	0.9633	1.0000	0.9813
NLPR	3	0.9510	1.0000	0.9749	0.9633	1.0000	0.9813
NLPR	4	0.9510	1.0000	0.9749	0.9633	1.0000	0.9813
NTU	1	0.9656	0.7693	0.8564	0.9748	0.7859	0.8702
NTU	2	0.9796	0.5798	0.7284	0.9878	0.5969	0.7441
NTU	3	0.9767	0.5796	0.7275	0.9866	0.5943	0.7418
TTRD	1	0.9507	0.6981	0.8051	0.9631	0.7006	0.8112
TTRD	2	0.9680	0.7363	0.8364	0.9759	0.7487	0.8474
*ISCAS	1	0.9703	0.9288	0.9491	0.9828	0.9369	0.9593

Table 5. Simplified Chinese relevance extraction results

Group	Run	Traditional				Simplified			
		Lenient		Strict		Lenient		Strict	
		T	RB	T	RB	T	RB	T	RB
CHUK	1	0.8254	0.3531	0.8238	0.2838	N/A			
iclpku	1	0.5872	0.2635	0.5797	0.1972	0.4124	0.2476	0.4275	0.1329
iclpku	2	0.5988	0.2490	0.5816	0.1688	0.4095	0.2353	0.1719	0.0480
*NTUCopeOpi	1	0.5028	0.2153	0.4825	0.1378	0.2909	0.1749	0.1792	0.1792
*NTUCopeOpi	2	0.4587	0.1939	0.4358	0.1225	0.0397	0.0239	0.0126	0.0126
*NTUCopeOpi	3	0.3191	0.1349	0.2969	0.0834	0.1587	0.0954	0.0776	0.0776
**TTRD	1	0.5645	0.2275	0.5496	0.1423	0.1129	0.0649	0.1719	0.0480
TTRD	2	0.5947	0.2365	0.5840	0.1565	0.1270	0.0711	0.1125	0.0345
NLPR	1	N/A				0.4286	0.2850	0.4759	0.2410
NLPR	2					0.4497	0.2389	0.4821	0.1971
NLPR	3					0.4037	0.2047	0.4689	0.1461
NLPR	4					0.4298	0.2451	0.4715	0.1965

Table 6. Performance (f-measure) of Chinese opinion holder task

5 Conclusion and Future Work

This paper introduces a Chinese opinion extraction system CopeOpi, which can extract relevant opinionated sentences, detect their polarities, and identify their holders. It is applied in NTCIR-7 MOAT task and performs the second in opinion extraction task and the third in polarity detection and sentence relevance judgment tasks. In this system, opinion scores are used to show the opinion polarities and strengths of words. The opinion scores of words are calculated not only considering their observation probability, but also their morphological types. Together with the negation operators and opinion operators, the opinion polarities can be determined for all sentences. The experimental results are satisfactory. We believe that we need to consider the sentence structures, at least the relation between words, if we want to improve the performance of opinion analysis.

The idea of distance is introduced in relevance sentence retrieval. We found that relevant sentences may cluster together, so we increased the relevance score of sentences close to the most relevant sentence. We also found that even though the term weight is very small, it still contributes to the relevance judgment. Reset the term weight smaller than the median does harm to the performance.

Though our system performs not well in the opinion holder extraction task, a preliminary approach has been proposed for further improvement. We extract opinion holders by a binary classifier trained lexical features and keywords. However, these features are not enough. There are still many issues to be discussed, including long opinion holders, translation problems, and common nouns as holders. To improve the performance of opinion holder extraction, we may need to extract more useful patterns and design more powerful rules. Also because we translated simplified Chinese corpus into a traditional Chinese one for processing it, and then translated the holders back into simplified Chinese to submit our runs, we should minimize the translation ambiguities to improve the performance.

NTCIR-6 Pilot task and NTCIR-7 MOAT task provided useful corpora for opinion analysis. Our future goal is to enhance our system by solving the problems we found this year by using available training instances.

References

[1] Cheng, X.-H. and Tian, X.-L. (1992). Modern Chinese. Bookman Books Ltd.

[2] Ku, L.-W., and Chen, H.-H. (2007). Mining Opinions from the Web: Beyond Relevance Retrieval. Journal of American Society for Information Science and Technology, Special Issue on Mining Web Resources for En-hancing Information Retrieval, 58(12), 1838-1850.

[3] Mierswa, Ingo, Wurst, Michael, Klinkenberg, Ralf, Scholz, Martin and Euler, Timm (2006). YALE: Rapid Prototyping for Complex Data Mining Tasks, Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[4] Kim, Youngho, Jung, Yuchul and Myaeng, S.-H. (2007). Identifying Opinion Holders in Opinion Text from Online Newspapers. Proceedings of the 2007 IEEE International Conference on Granular Computing.

[5] Tsuchiya, Masatoshi, Hida, Shinya, Nakagawa, Seiichi (2008). Robust Extraction of Named Entity Including Unfamiliar Word, Proceedings of ACL-08: HLT, Short Papers (Companion Volume), pages 125–128, Columbus, Ohio, USA.

[6] Pang, Bo and Lee, Lillian (2008). Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval, Vol. 2, No 1-2 (2008) 1–135.

[7] Bautin, Mikhail, Vijayarenu, Lohit and Steven Skiena (2008). International sentiment analysis for news and blogs. In Proceedings of the International Conference on Weblogs and Social Media (ICWSM).

[8] Yohei Seki, David Kirk Evans, Lun-Wei Ku, Le Sun, Hsin-Hsi Chen, Noriko Kando (2008). Overview of Multilingual Opinion Analysis Task at NTCIR-7. In Proceedings. of the Seventh NTCIR Workshop.