

Identification of Opinion Holders

Lun-Wei Ku*, Chia-Ying Lee* and Hsin-Hsi Chen*

Abstract

Opinion holder identification aims to extract entities that express opinions in sentences. In this paper, opinion holder identification is divided into two subtasks: author's opinion recognition and opinion holder labeling. Support vector machine (SVM) is adopted to recognize author's opinions, and conditional random field algorithm (CRF) is utilized to label opinion holders. New features are proposed for both methods. Our method achieves an f-score of 0.734 in the NTCIR7 MOAT task on the Traditional Chinese side, which is the best performance among results of machine learning methods proposed by participants, and also it is close to the best performance of this task. In addition, inconsistent annotations of opinion holders are analyzed, along with the best way to utilize the training instances with inconsistent annotations being proposed.

Keywords: Opinion Holder Identification, Opinion Mining, Conditional Random Field, Support Vector Machine.

1. Introduction

Opinions describe subjective thinking of people. With the blooming of Web 2.0, a large number of free and online articles have become easily accessible. Although people are interested in the shifting of opinions, they cannot read such a large quantity of articles in a short time. Opinion mining can analyze opinions from many information sources automatically and helps extract opinions, along with determining their polarities, strength, holders, and targets. Opinion polarities tell us whether the current opinions are positive, neutral, or negative. The opinion strength then tells us the degree of their attitude, *i.e.*, strong, medium, or weak. Opinion holders are the people who express opinions, while opinion targets are the objects of those opinions. Let us take "Mr. Wang loves to play baseball" as an example. In this opinion sentence, its polarity is positive, its strength is strong, the opinion holder is Mr. Wang, and the opinion target is playing baseball. It is an opinion from Mr. Wang that indicates

* Department of Computer Science and Information Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, 10617 Taiwan, TEL: +886-2-33664888*311, FAX: +886-2-23628167
E-mail: {lwku, cylee}@nlg.csie.ntu.edu.tw; hhchen@csie.ntu.edu.tw

he has a positive attitude towards playing baseball.

Opinion holder identification is useful in knowing who has the same attitude, what kind of issues a specific person cares about, and whether there are different opinions from some specific persons. This technique can also be applied to social network analysis to discover who the opinion leader is. It is also important in an opinion question answering system as it can provide the owner of opinions.

There are three major challenges in opinion holder identification: co-reference resolution, parsing nested opinions, and inconsistent annotation utilization. Like the conventional question answering problem, pronoun-antecedent and zero anaphor problems have to be solved before identifying opinion holders. Nested opinions are common in long sentences. People like to quote opinions of other people to show that they are impartial, but this behavior also implies that they agree with their quotes. In this case, we need to identify both the quoting and the quoted holders for further analysis.

It is sometimes difficult to determine the holder of an opinion. For example, even though the holder is obvious in the opinion sentence, we may find that he represents some organization and is presenting the organization's opinion. The following sentence is an example: "According to the media, [the] U.S. and China are discussing the agreement of terminating the usage of nuclear weapons; Becon said they have discussed this issue before." In this sentence, "they" refers to the U.S. and China. Becon quoted words from the U.S. and China, so this is a nested structure. In addition, although this expression is said by the media and Becon, the holder should be the U.S. and China. These challenges all complicate the annotation process, and a double check and a selection process are necessary when generating the gold standard.

2. Related Work

Pang and Lee (2008) have mentioned some important research projects in the domain of opinion mining. Kim and Hovy (2004) proposed four elements in opinion mining, including the opinion polarity, the opinion strength, the opinion holder, and the opinion target. Among them, the research for opinion holder identification is new. Previous researchers mainly have proposed two kinds of methods: heuristic rule based and machine learning based methods.

2.1 Heuristic Rule based Methods

For heuristic rule based methods, Seki *et al.* (2009) utilized noun phrases and linguistic features, and adopted SVM to classify opinion holders into authors and non-authors in English and Japanese materials. Xu and Wang (2008) first solved the co-reference resolution, and extracted opinion holders by rules involving punctuation marks, conjunctions, prefixes,

suffixes, and opinion operators. They achieved an f-score of 0.825 in the NTCIR7 MOAT task on the Traditional Chinese side, which is the state of the art.

2.2 Machine Learning based Methods

For the machine learning based methods, many researchers have adopted maximum entropy algorithms, SVM, or the conditional random field model. Kim and Hovy (2006) utilized the maximum entropy model to extract opinion holders and targets from news articles. They first found opinion words and labeled semantic roles, then identified the semantic roles that are opinion holders and targets. Kim (2007, 2008) classified opinion holders into authors, simple holders and co-referenced holders, then extracted lexical and syntactic features for SVM to select the best opinion holder. So far, this is the best method for English materials, and it achieved an f-score of 0.346. Wu (2008) used words and parts of speech as features in L2-norm linear SVM to solve this research problem as a similar method for named entity identification. Breck and Choi (2007, 2005) utilized lexical features, syntactic features, dictionary-based features, and dependency features by CRF to identify opinion holders. Meng and Wang (2008) used words, parts of speech, and opinion operators, while Liu and Zhao (2008) extracted parts of speech, semantic features, contextual features, dependency features, and position features by CRF.

2.3 Proposed Methods

We propose a unique approach that divides opinion holder identification into two tasks: author's opinion recognition and opinion holder labeling. We then find better strategies to perform these two tasks. For author's opinion recognition, we adopt SVM by features such as words and their parts of speech, named entities, punctuation marks, the context, and opinion related information in the current sentence. Among them, some context features (the roles of verbs) and opinion related features (information of positive words, neutral words, negative words, and opinion operators) have not been utilized in opinion holder identification before. Detailed features will be described in Sections 3.2 and 3.3.

3. Opinion Holder Identification

Five procedures of opinion holder identification are proposed in this paper, including text pre-processing, author's opinion recognition, opinion holder labeling, post-processing, and result generation. Chinese word segmentation, parts of speech tagging, and named entity recognition are performed in the text pre-processing stage. Then, author's opinions are recognized and opinion holder labeling determines the text segment referring to the holder. We have two strategies for applying the proposed methods of author's opinion recognition and opinion holder labeling. These two strategies are described in Section 3.5. After that, this

labeled text segment is processed by the post-processing procedure to generate the final opinion holder. The flowchart describing these five procedures is shown in Figure 1.

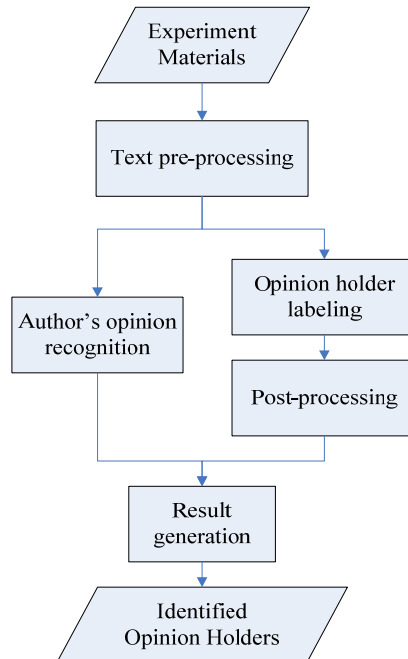


Figure 1. Flowchart of opinion holder identification

3.1 Text Pre-processing Stage

In the text pre-processing stage, we utilize the Chinese word segmentation system developed by Lo (2008). We, however, modify its segmentation module and add additional name dictionaries to it. The length limit of the modified Chinese name module is set looser and Japanese family names are added so that the segmentation system can recognize Japanese names, which are usually longer than Chinese names. Occupations, titles, and company names are also added to the dictionary of the segmentation system to provide useful holder relevant information.

3.2 Author's Opinion Recognition Stage

Author's opinion recognition finds out whether the opinion holder of the current sentence is the author. In this paper, this task is viewed as a binary classification task and LIBSVM (Chang & Lin, 2001) is adopted for classification. The main features extracted for this task are words, parts of speech, named entities, punctuation marks, sentence components, and opinion operators. Table 1 shows all of the features utilized in this task. The lexicon features include

first person pronouns, which are often utilized by the authors to refer to themselves. The part-of-speech features include general pronouns and personal pronouns, because pronouns usually refer to persons and organizations and they can express opinions. The named-entity features are considered because they can either be the opinion holders or the opinion targets.

Table 1. Features for author's opinion recognition

Feature Type	Feature Name	Feature Description
Lexicon	FHasI	Does the word “I” (我) appear?
	FHasWe	Does the word “we” (我們) appear?
	fNumI	The number of the word “I” (我)
	fNumWe	The number of the word “we” (我們)
Part of speech	fHasPronoun	Are there any pronouns?
	fHasManPronoun	Are there any personal pronouns?
	fNumPronoun	The number of pronouns
	fNumManPronoun	The number of personal pronouns
Named entity	fHasPer	Is there a person name (named entity)?
	fHasLoc	Is there a location name (named entity)?
	fHasOrg	Is there an organization name (named entity)?
	fHasNa	Are there any common nouns?
	fHasNb	Are there any proper nouns?
	fHasNc	Are there any common location nouns?
	fNumLoc	The number of location names (named entity)
	fNumOrg	The number of organization names (named entity)
	fNumPer	The number of personal names (named entity)
	fNumNa	The number of common names
	fNumNb	The number of proper names
fNumNc	The number of common location names	
Punctuation mark	fHasExclamation	Is there an exclamation mark (“!” or “!”) ?
	fHasQuestion	Is there a question mark (“?” or “?”) ?
	fHasColon	Is there a colon (“:” or “:”) ?
	fHasLeftQuotation	Are there any quotation marks (“ ” or “【”)?
	fHasRightQuotation	Are there any quotation marks (“ ” or “】”)?
Sentential	fNumChar	The number of Chinese characters
	fNumWord	The number of Chinese words
	fNumSubsen	The number of clauses
Opinion	fOperator1 to 203	Is there an opinion operator?

The punctuation features include punctuation marks that possibly co-occur with opinions. For example, exclamation points and question marks often appear in sentences expressed by people because they can bear sentiment, whereas colons and quotations are usually used to quote expressed words. The sentential features tell the length of the sentences by their composite characters, words, and clauses. We consider these features because we think that authors may need a sentence of a suitable length to express opinions. As to the opinion features, a total of 203 opinion operators, such as 說 (say), 指出 (point out), and 認為 (think), are collected manually from the earlier NTCIR corpus (Seki *et al.*, 2008) for this task.

3.3 Opinion Holder Labeling Stage

Opinion holder labeling finds the text segment that represents the opinion holder. In the beginning, this task is also viewed as a binary classification problem for all words of a sentence, where the decision tree determines whether the current word is part of the opinion holder or not. CHAID decision tree algorithm provided by RapidMiner (Mierswa, Wurst, Klinkenberg, Scholz, & Euler, 2006) is adopted. It is a pruned decision tree using the chi-square test.

As the other alternative, we view the opinion holder labeling problem as a sequential labeling problem. Therefore, the CRF algorithm (Lafferty, McCallum, & Pereira, 2001) is selected to label whether each composite word is a portion of the opinion holder and CRF++ (Kudo, 2003) is adopted for implementation. Features for experiments are listed in Table 2.

Table 2. Features for opinion holder labeling

Feature Type	Feature Name	Feature Description
Lexicon	fWord	The current Word
Part of speech	fPOS	Part of speech of the current word
	fIsPronoun	Is the current word a pronoun?
	fIsNoun	Is the current word a noun?
Named entity	fIsPer	Is the current word a person name?
	fIsLoc	Is the current word a location name?
	fIsOrg	Is the current word an organization name?
Punctuation mark	fAfterParen	Does the current word appear one word after a parenthesis “ ┘ ” or “ ┐ ”?
	fBeforeColon	Does the current word appear one word before a colon “ : ” or “ ⋮ ”?
Sentential	fNearSenStart	Is the current word close to the sentence head?
	fSenLen	The number of words in the current sentence
	fWordOrder	The absolute position of the current word in the sentence
	fWordPerc	The absolute position (in percentage) of the current word in the sentence

Context	fNearVerb	The nearest verb in the current sentence
	fNearVerbPOS	The type (POS) of the nearest verb ¹ , <i>e.g.</i> , VA (transitive verb), VB (intransitive verb), <i>etc.</i>
	fDistNearVerb	The distance between the current word and its nearest verb
Opinion	fHasOpKW	Is there an opinion operator in this sentence?
	fHasPosKW	Are there any positive opinion words in this sentence, <i>e.g.</i> , 成功 (success), 同意 (agree)?
	fHasNegKW	Are there any negative opinion words in this sentence, <i>e.g.</i> , 錯誤 (wrong), 失敗 (fail)?
	fHasNeuKW	Are there any neutral opinion words in this sentence, <i>e.g.</i> , 不予置評 (no comment), 兩難 (a difficult choice)?
	fNearOpKW	Nearest opinion word in this sentence
	fNearPosKW	Nearest positive opinion word in this sentence
	fNearNegKW	Nearest negative opinion word in this sentence
	fNearNeuKW	Nearest neutral opinion word in this sentence
	fNearOpKWPOS	POS of the nearest opinion operator
	fNearPosKWPOS	POS of the nearest positive opinion word
	fNearNegKWPOS	POS of the nearest negative opinion word
	fNearNeuKWPOS	POS of the nearest neutral opinion word
	fDistOpKW	The distance to the nearest opinion operator in this sentence
	fDistPosKW	POS of the nearest NTUSD positive opinion word in this sentence
	fDistNegKW	POS of the nearest NTUSD negative opinion word in this sentence
fDistNeuKW	POS of the nearest NTUSD neutral opinion word in this sentence	

Features for opinion holder labeling include words, parts of speech, named entities, punctuation marks, sentential information, contextual information, and opinion related information. Some of the features are the same as those we have selected for the author's opinion recognition. The lexicon feature is the current word to be determined. The part-of-speech features of the current word include its part of speech, and whether it is a noun or a pronoun. The binary properties of being a noun or a pronoun are emphasized here because they are the most commonly seen parts of speech in opinion holders. Punctuation marks also are considered as features here. Sentential features tell the position of the current word in the current sentence. They are included in the feature set because, according to our observations,

¹ The part of speech tagging set is listed in Technical Report no. 95-02/98-04, Chinese Knowledge Information Processing Group, Academia Sinica.

holders often appear in the beginning or at the end of the sentence. The context features include the information of the nearest verbs with respect to the current word. If the current word is a part of the opinion holder, its nearest verb could be an opinion operator. The collocation of the current word and the nearest verb are considered. For the opinion information, the appearance of opinion operators, positive words, neutral words, and negative words are considered. Here, positive words are words used to express a supportive attitude, such as success, good, *etc.*; neutral words express an impartial attitude, such as no comment, difficult to say, *etc.*; negative words express opposite attitude, such as objection, accusation, *etc.* The occurrence of opinion words may indicate the existence of opinions, and opinions may further indicate the existence of their holders. The method of utilizing contextual and opinion information, along with the features of nouns and pronouns, are first proposed in this paper.

The training set for the NTCIR-7 MOAT Task, introduced in Section 4.1, is adopted for extracting training features and for building models. As the size of this set is not large, the co-training method is adopted to improve the performance (Blum & Mitchell, 1998). Co-training is a semi-supervised learning method that trains models together with labeled and un-labeled materials. In co-training, sentences with high CRF confidence scores are selected, and sentences among them without words that are portions of the opinion holder are dropped. These selected sentences, along with their predicted labels, are fed back to the CRF system as the training sentences in the next iteration. The co-training process is shown in Figure 2.

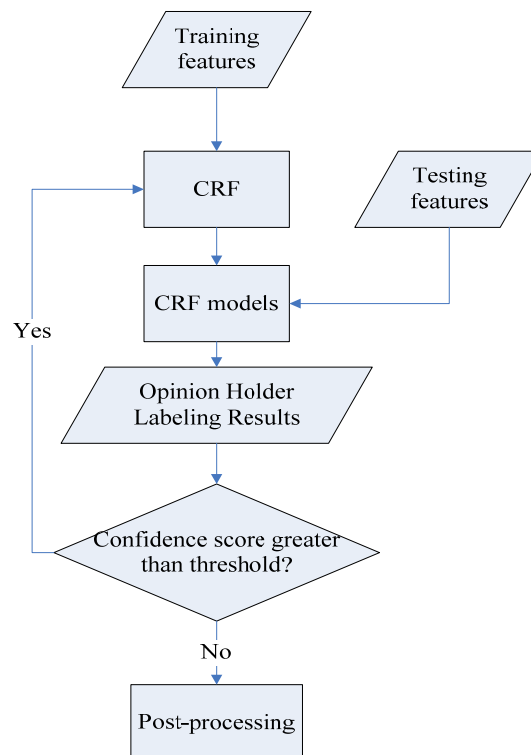


Figure 2. Co-training process

3.4 Post-processing Stage

Post-processing includes two steps: processing phrasal opinion holders and recovering named entities. As mentioned above, opinion holder labeling tells whether the current word is a portion of the opinion holder. Nevertheless, opinion holders often are composed of multiple words, and we need to combine these words to propose the final result if the holder is longer than one word.

3.4.1 Processing Phrasal Opinion Holders

In our experiments for processing phrasal opinion holders, five labels are used to label words in CRF opinion holder labeling: H (head word), I (middle word), T (tail word), S (single word), and O (not opinion portion), abbreviated as HITSO hereafter. Instead, when working with CHAID, the decision tree, it can only generate two labels “YES” and “NO” to tell whether the current word is a portion of the opinion holder. In this phase, comparing the performances of using different label sets is not the focus, so we just use the label set HITSO. The effect of different labels will be tested later.

Different post-processing rules are applied according to the tagging sets. If words are labeled by the tagging set H, I, T, S, and O by CRF, we use the following rules to combine words to a phrase if necessary:

- (1) Find the H labeled word with the highest confidence score in the current sentence.
- (2) Combine the sequent I labeled words until a T labeled word is found.
- (3) The final opinion holder includes words that begin from the word with label H found in (1) to the word with label T found in (2).
- (4) If all words in the current sentence are all labeled “O”, the opinion holder will be set to the author.

If words are labeled with the tagging set {YES, NO} by CHAID, we use the following three rules to combine words to a phrase if necessary:

- (1) Combine consecutive nouns. For example, “印度 (India, Nc) 總統 (president, Na) 瓦希德 (Abdurrahman Wahid, Nb)” are combined into one opinion holder phrase.
- (2) Use conjunctions and “、” (a mark in Chinese punctuations to list items in a series) to combine nouns into one holder group. For example, “瓦希德 (Abdurrahman Wahid, Nb) 、 (PAUSECATEGORY) 柯林頓 (Clinton, Nb) 與 (and, Caa) 小淵惠三 (Keizo Obuchi, Nb)” are combined into one opinion holder group. If people express the same opinion together, they are usually grouped as an opinion holder group in sentences.
- (3) If all words in the current sentence are labeled “NO,” the opinion holder will be set to the author.

3.4.2 Recovering Named Entities

Translated named entities tend to be segmented incorrectly in text pre-processing. These segmentation errors will make opinion holders incomplete, so recovering wrongly segmented named entities is necessary. In this paper, we postulate that the number of occurrences of a complete named entity should be the same as its partial sequences. Therefore, we will compare the occurrences of the current holder sequence with the current holder sequence plus its previous/following word to decide whether we should combine them to generate a more complete opinion holder. For example, in the sentence “Indonesian big man Suharto ruled Indonesia for 32 years,” the name “Suharto” is translated into “蘇哈托”. Nevertheless, this name is wrongly segmented into two words ”蘇哈” and “托”. In this case, we will check whether the number of appearance of “蘇哈托” is the same as ”蘇哈”. If it is, we will combine ”蘇哈” and “托” into one word “蘇哈托”. This process is done iteratively until the numbers of appearance of the current word and that plus the previous/following word are not the same. In this example, we further check the next word “統”. In doing so, we will find that the numbers of appearance of “蘇哈托” and “蘇哈托統” are not the same. Therefore, the recovery process stops and we propose “蘇哈托” as the opinion holder. This process is shown as follows.

```

For each sentence  $S = \{w_1, w_{i+1}, \dots, w_n\}$ ,  $w$  is a word in  $S$ 
where the identified opinion holder  $OP = \{w_i, w_{i+1}, \dots, w_j\}, i \geq 1, j \leq n$  :
while  $i > 1$ 
    if the number of occurrence of  $OP$  equals that of the string  $\{w_{i-1}, OP\}$ ,  $i--$ ;
while  $j < n$ 
    if the number of occurrence of  $OP$  equals that of the string  $\{OP, w_{j+1}\}$ ,  $j++$ ;

```

3.5 Result Generation Stage

The final step is result generation, which considers the results of the author’s opinion recognition and opinion holder labeling. Author’s opinion recognition classifies all opinion sentences into author’s opinions and non-author’s opinions, while opinion holder labeling labels the opinion holder. We propose two result generation strategies, and their flowcharts are shown in Figures 3 and 4.

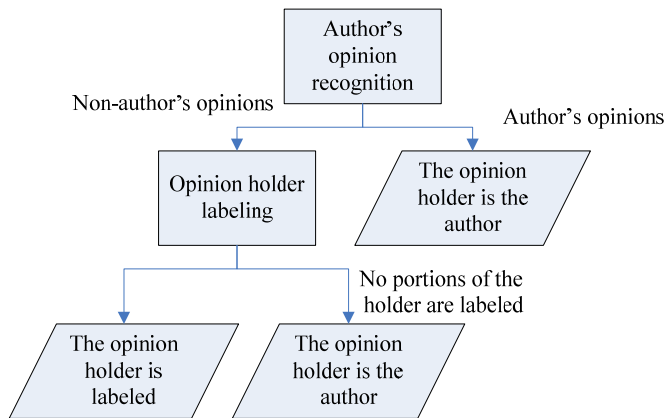


Figure 3. Result generation strategy A

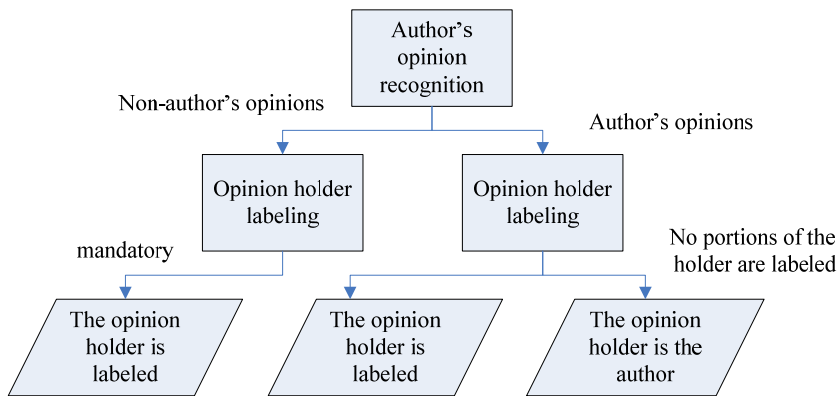


Figure 4. Result generation strategy B

In Result generation strategy A, we have more confidence in the author’s opinions recognized by the author’s opinion recognition module. Non-author’s opinions are passed to the opinion holder labeling module. If there is an opinion holder labeled by this module, we propose it; otherwise, we propose the author as the opinion holder. In Result generation strategy B, we have more confidence in the non-author’s opinions recognized by the author’s opinion recognition module. Both kinds of opinions are then passed to the opinion holder labeling module. For the non-author’s opinions, we force the opinion holder labeling module to propose an opinion holder by considering the most possible opinion holder among words in the current sentence. For the author’s opinions, if there is an opinion holder labeled by this module, we propose it; otherwise, we propose the author as the opinion holder. After the text pre-processing, author’s opinion recognition, opinion holder labeling, post-processing, and the result generation, the opinion holder is determined.

4. Experiment and Discussion

In this section, the experimental corpus and resources are introduced. Results of author's opinion recognition, opinion holder labeling, and the complete opinion holder identification are shown and discussed.

4.1 Corpus and Evaluation Tasks

The adopted experimental corpus is the NTCIR-6 Pilot Task & NTCIR-7 MOAT Task Traditional Chinese corpora. NTCIR² is one of the three important international evaluative forums. MOAT (Multilingual Opinion Analysis Task) is one of its evaluative tasks (Seki *et al.* 2008). The MOAT task provides English, Japanese, Traditional Chinese, and Simplified Chinese materials, which include news articles collected from 1998 to 2001. Labels for relevance, the opinion sentence, the opinion polarity, and the opinion holder are provided by the NTCIR-6 Pilot Task. In addition to these labels, labels for the opinion target are provided later by the NTCIR-7 MOAT Task.

Each MOAT corpus is classified into training and testing sets. The original purpose of the corpus released prior to the formal run (the training corpus mentioned here) is to provide samples to the participants, so its quantity is comparably small. The NTCIR-7 MOAT training set includes documents of 3 topics, consisting of 1,509 sentences, with 944 of them being opinion sentences; the NTCIR-7 MOAT testing set includes documents of 14 topics, consisting of 4,665 sentences, with 2,174 of them being opinion sentences. The opinion labels are sentence-based. As the size of the NTCIR-7 MOAT training set is small, the NTCIR-6 Pilot Task testing set is added for training in this paper. This set includes documents of 29 topics, consisting of 9,240 sentences, with 5,453 of them being opinion sentences. Labels for opinion sentences and opinion holders in these training and testing sets are utilized for opinion holder identification in this paper.

4.2 Experimental Resources

The opinion dictionary and named entity dictionaries are adopted in this paper. The opinion dictionary includes opinion operators, positive, neutral, and negative opinion words extracted from the NTCIR7 MOAT training set, and NTUSD (Ku & Chen, 2007). It is utilized for feature extraction. The person name, location name, and organization name dictionaries are collected for named entity recognition here, including the Million person name dictionary, Sinica corpus³, CNA translated name dictionary, The Revised Chinese Dictionary⁴, Japanese

² <http://research.nii.ac.jp/ntcir/>

³ <http://dbo.sinica.edu.tw/ftms-bin/kiwi1/mkiwi.sh?language=1>

⁴ <http://dict.revised.moe.edu.tw/>

common family name dictionary, MOE translated location name dictionary, translated foreign location name list, and Taiwan national industry list.

4.3 Author’s Opinion Recognition

In this experiment, the NTCIR-6 testing set and the NTCIR-7 training set were utilized for training, while the NTCIR-7 testing set was used for testing. NTCIR generates the gold standard under two metrics: the strict metric and the lenient metric. We selected opinion sentences by the lenient metric as the gold standard for testing, *i.e.* for these sentences, at least two out of three annotators label them as opinions. Precision, recall, f-score, and accuracy were adopted for evaluation.

The experimental corpus was annotated by three annotators. Therefore, there are inconsistent annotations of opinion holders in some opinion sentences. For example, in the sentence “Because Taiwan understands, using nuclear weapons will destroy the relationship with the U.S.” one annotator labeled “Taiwan” as the holder, while the other selected “the author”. The example sentence is an implicit nested opinion, and inconsistent annotations are found often in nested opinions. We checked all of the sentences to see if there are many inconsistently labeled opinion holders. The results are shown in Figure 5.

For the NTCIR-6 corpus, if the annotators could find any opinion holder, it was reported; if they could not, the opinion holder was automatically set to the author. Therefore, we cannot know how many opinion holders are inconsistently labeled (see the “?” in Figure 5). Instead, for the NTCIR-7 corpus, the opinion holder “the author” was explicitly annotated. Therefore, we are able to find the percentage of inconsistently labeled opinion holders. Figure 5 shows that the opinion holder in 19% of sentences is consistently labeled as the author, while that in 15% of sentences is inconsistently labeled as the author and the other named entity. These two percentages are close to each other.

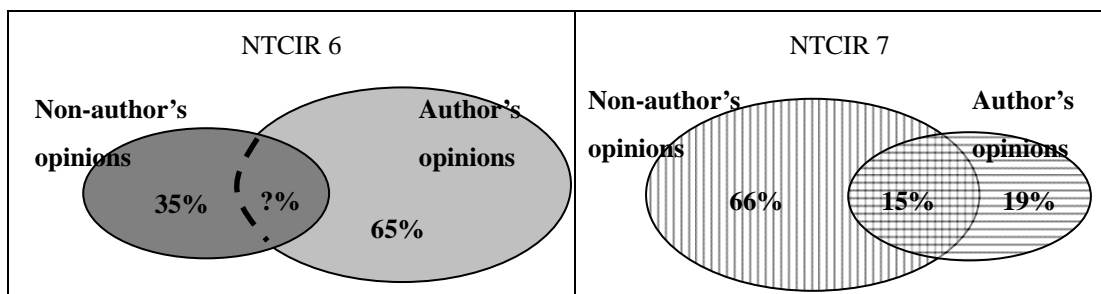


Figure 5. The percentages of the author’s opinions in NTCIR-6 and NTCIR-7 corpora

From experiments, we have found that using opinion sentences as the training materials performs better than using all sentences and using both NTCIR-6 (all) and NTCIR-7 (testing)

sentences for training also performs better than only using one of them. Therefore, in the experiments for author's opinion recognition, opinion sentences of NTCIR-6 and NTICR-7 were both used for training.

We have discussed the inconsistency in the gold standard for the author's opinion recognition. Therefore, three settings are tested: treating inconsistently labeled opinions as the author's opinions, treating inconsistently labeled opinions as non-author's opinions, and expelling inconsistent labeled opinions in the training set. Table 3 shows the testing performances of these three settings. From Table 3, we find that the first setting, treating inconsistently labeled opinions as the author's opinions, performs the best. It achieves an f-score of 79.98%, which outperforms the setting of treating them as non-author's opinions. The f-score, 65.0%, is even worse when the inconsistently labeled data is not considered, compared to the former two settings. Therefore, we conclude that inconsistently labeled data is useful and we should treat them as the author's opinions.

Table 3. Performances of three settings for inconsistently labeled author's opinions

Settings	Precision	Recall	f-score	Accuracy
Author's opinions	69.68%	93.85%	79.98%	83.49%
Non-author's opinions	64.87%	95.94%	77.40%	80.31%
Expelling inconsistency	50.52%	91.53%	65.10%	77.28%

4.4 Opinion Holder Labeling

For opinion holder labeling, the NTCIR-7 training set is utilized. Here, we compare the performances of using strict opinion sentences (agreed on by three annotators) and lenient opinion sentences for training. After labeling words in sentences, the results of author's opinion recognition and opinion holder labeling are considered together to generate the proposed opinion holder. In this experiment, only opinion sentences correctly proposed by the system are evaluated, so the real performance of the opinion holder labeling can be calculated without the propagation errors from opinion extraction. As the number of sentences is the same as the number of opinion holders, the precision, recall, and f-score are not used as evaluation metrics because they will be equal. Instead, the number of correct holders and wrong holders, and the set f-score is adopted. The formula for calculating the set f-score is shown below.

$$set - f - score = \frac{system_correct(holder)}{system_correct(opinion = yes)} \quad (1)$$

As mentioned in the previous section, CHAID and CRF are both tested for their performances in opinion holder labeling. Their performances are shown in Table 4.

Table 4. Performances of opinion holder labeling using CHAID and CRF

	Method	Correct #	Error #	Set f-score
Strict	CHAID	564	605	48.16%
	CRF	818	351	69.89%
	CHAID+CRF	825	344	70.57%
Lenient	CHAID	981	967	50.31%
	CRF	1317	631	67.57%
	CHAID+CRF	1322	626	67.83%

Table 4 shows that CRF performs much better than CHAID in opinion holder labeling for both strict and lenient opinion sentences. In the setting CHAID+CRF, we first use CHAID to get the predicted labels, and use these labels together with other features as the input features for CRF. Results show that this setting can slightly improve the performance. One reason could be that the tagging set for CRF is larger than CHAID. Another reason could be that CRF has better performance in combining words with labels, while CHAID needs to apply rules on the proposed labels to find the complete opinion holder. The best performances achieved are a set f-score of 70.57% for strict opinion sentences and 67.83% for lenient opinion sentences.

Furthermore, we classify labeling errors into six types according to the position of the proposed opinion holder and the correct opinion holder:

- (1) The proposed opinion holder is not related to the correct holder in the aspect of position. 29.1% errors are of this type.
- (2) The proposed opinion holder has one additional word in the front or rear, compared to the correct opinion holder. For example, “魯斯曼日前” (Russman the other day, proposed) and “魯斯曼” (Russman, correct). 18.1% errors are of this type.
- (3) The system only proposes the title of the opinion holder, but not the name. For example, “科索伏著名塞裔領袖” (The famous Serbian leader of Kosovo, proposed) and “科索伏著名塞裔領袖特拉伊科維契” (The famous Serbian leader of Kosovo Trajkovic, correct). 8.3% errors are of this type.
- (4) The system only proposes the modifier of the correct opinion holder. For example, “該” (That, proposed) and “該裁決” (That decide, correct). 7.5% errors are of this type.
- (5) The proposed opinion holder has two or more additional words compared to the correct opinion holder. For example, “狄蘭在記者會” (Dylan proposed in the press conference) and “狄蘭” (Dylan, correct). 5.5% errors are of this type.
- (6) The proposed opinion holder is incomplete. 4.7% errors are of this type.

From these errors, we find that most errors occur because the system cannot determine

the first word and the last word of the opinion holder properly. Therefore, different tagging sets were tested (IO, ISO, HTO, HISO, HITSO, etc.), and we have found that using the tagging set {H, I, O} can achieve the best performance, which is the set f-score of 70.57% for strict opinion sentences. We also propose the name entity recovery method to deal with errors of the sixth type. Experiments show that with co-training, the best confidence score threshold 0.7 and the named entity recovery, our system achieves the best set f-score of 72.03%.

4.5 Experiments for Opinion Holder Identification

After author's opinion recognition and opinion holder labeling, we need to propose the opinion holder according to these results. Table 5 shows the performances of applying different result generation strategies.

Table 5. The performance of opinion holder identification: applying different result generation strategies

	Strategy	Correct #	Wrong #	Set f-score
strict	A	829	340	70.92%
	B	858	310	73.40%
lenient	A	1338	611	68.65%
	B	1372	576	70.40%

Table 5 shows that the performance of applying Result generation strategy B is better than applying Result generation strategy A. It indicates that the proposed method for author's opinion recognition works better in determining non-author's opinion. That is, we can be more sure when the system tells that the current opinion is not expressed by the author. Using fewer author relevant features may be the reason for this phenomenon.

4.6 Performances of NTCIR-7 Participants

NTCIR-7 evaluates the system performance in two ways. One is to evaluate sentences correctly proposed by the system, which is also used in the previous section; the other is to evaluate all opinion sentences in the testing set.

Table 6 shows the performances of all participants' systems together with the performance of our system. WIA's system performs the best in both evaluation metrics. WIA adopts heuristic rules to design their systems. Therefore, our system performs the best among all systems using machine learning methods. Moreover, our system performs better for strict opinion sentences, which is different from other systems. In other words, our system is good at identifying the holder of opinions that were consistently annotated by annotators.

Table 6. Performances of NTCIR-7 participants

	Participants	Number of correctly proposed opinion sentences	Evaluate correctly proposed opinion sentences	Evaluate all opinion sentences		
			Set f-Score	Precision	Recall	f-score
Strict	WIA	757	82.30%	19.88%	49.52%	28.38%
	iclpku-1	880	57.84%	13.03%	40.53%	19.72%
	iclpku -2	989	58.04%	10.35%	45.70%	16.88%
	TTRD-1	1213	54.91%	8.22%	52.95%	14.23%
	TTRD-2	866	58.31%	9.72%	40.13%	15.65%
	NTU-1	1169	48.16%	8.14%	44.90%	13.78%
	Our System	1169	73.40%	12.38%	68.31%	20.97%
Lenient	WIA	1134	82.54%	29.92%	43.05%	35.31%
	iclpku-1	1364	58.72%	20.51%	36.84%	26.35%
	iclpku -2	1606	59.90%	17.33%	44.20%	24.90%
	TTRD-1	2070	56.47%	16.78%	40.02%	23.65%
	TTRD-2	1464	59.49%	14.43%	53.73%	22.75%
	NTU-1	1948	50.31%	14.43%	53.73%	22.75%
	Our System	1948	70.40%	19.80%	63.11%	30.15%

5. Conclusion and Future Work

We proposed a machine learning based method for opinion holder identification. We classified this task into two subtasks: author's opinion recognition and opinion holder labeling. SVM was adopted for author's opinion recognition, and CRF was adopted for opinion holder labeling. We proposed lexical, syntactic, contextual, and opinion features. Named entities and punctuation marks were also utilized as features. We tested different tagging sets to find the best set {H, I, O}. Co-training was proposed to solve the problem of insufficient training materials, and results merging strategies were proposed to improve the performance. We also mentioned the methods of utilizing inconsistent annotated materials and analyzed system errors to find solutions for improving the performance.

The proposed system for the opinion holder identification achieved an f-score of 0.734, which is the best among machine learning based systems and is close to the state of the art. The state of the art system adopts heuristic rules. Nevertheless, heuristic rule based systems like it are difficult to rebuild because the rules are usually not described in detail in the previous literature.

In the future, we hope to solve the co-reference resolution problem, which is important in

named entity extraction and also in opinion holder extraction. In addition, we plan to add parsing information to improve the performance. Finding a good named entity recovery algorithm is also one of our next attempts. In summary, utilizing techniques of opinion holder identification in the opinion analysis system to compare opinions of different people is our future goal.

References

- Blum, A., & Mitchell, T. (1998). Combining labeled and unlabeled data with co-training. *Conference on Computational Learning Theory*, 92-100.
- Breck, E., Choi, Y., & Cardie, C. (2007). Identifying expressions of opinion in context. *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 2683-2688.
- Chang, C. C., & Lin, C. J. (2001). LIBSVM: a library for support vector machines.
- Choi, Y., Cardie, C., Riloff, E., & Patwardhan, S. (2005). Identifying sources of opinions with conditional random fields and extraction patterns. *Proceedings of EMNLP conference*, 355-362.
- Kim, S. M., & Hovy, E. (2004). Determining the sentiment of opinions. *Proceedings of the COLING conference*, 1367-1374.
- Kim, S. M., & Hovy, E. (2006). Extracting opinions, opinion holders, and topics expressed in online news media text. *Proceedings of the Workshop on Sentiment and Subjectivity in Text at the joint COLING-ACL conference*, 1-8.
- Kim, Y., Jung, Y., & Myaeng, S. H. (2007). Identifying opinion holders in opinion text from online newspapers. *International Conference on Granular Computing*, 699-702.
- Kim, Y., Kim, S., & Myaeng, S. H. (2008). Extracting topic-related opinions and their targets in NTCIR-7. *Proceedings of the Seventh NTCIR Workshop*, 247-254.
- Ku, L. W., & Chen, H. H. (2007). Mining opinions from the web: beyond relevance retrieval. *Journal of American Society for Information Science and Technology*, 1838-1850.
- Kudo, T. (2003). CRF++: yet another CRF toolkit. <http://crfpp.sourceforge.net/>.
- Lafferty, J., McCallum, A., & Pereira, F. (2001). Conditional random fields: probabilistic models for segmenting and labeling sequence data. *Proceedings of International Conference on Machine Learning*, 282-289.
- Liu, K., & Zhao, J. (2008). NLPR at Multilingual Opinion Analysis Task in NTCIR7. *Proceedings of the Seventh NTCIR Workshop*, 226-231.
- Lo, Y. S. (2008). *An Approach of Using Multiple Dictionaries and Conditional Random Field in Chinese Segmentation and Part of Speech Tagging*. Master Thesis, National Taiwan University.
- Meng, X., & Wang, H. (2008). Detecting opinionated sentences by extracting context information. *Proceedings of the Seventh NTCIR Workshop*, 268-271.

- Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006). YALE: rapid prototyping for complex data mining tasks. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 935-940.
- Pang, B., & Lee, L. (2008). *Opinion mining and sentiment analysis*. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.
- Seki, Y., Evans, D. K., Ku, L. W., Sun, L., Chen, H. H., & Kando, N. (2008). Overview of multilingual opinion analysis task at NTCIR-7. *Proceedings of the Seventh NTCIR Workshop*, 185-203.
- Seki, Y., Kando, N., & Aono, M. (2009). Multilingual opinion holder identification using author and authority viewpoints. *Journal of Information Processing and Management*, 189-199.
- Wu, Y. C., Yang, L. W., Shen, J. Y., Chen, L. Y., & Wu, S. T. (2008). Tornado in multilingual opinion analysis: a transductive learning approach for Chinese sentimental polarity recognition. *Proceedings of the Seventh NTCIR Workshop*, 301-306.
- Xu, R., & Wong, K. F. (2008). Coarse-Fine opinion mining - WIA in NTCIR-7 MOAT task. *Proceedings of the Seventh NTCIR Workshop*, 307-313.

