Incorporating Feature-based and Similarity-based Opinion Mining – CTL in NTCIR-8 MOAT

Ruifeng Xu
[1]Dept. of Chinese, Translation and Linguistics,
City University of Hong Kong
Kowloon, Hong Kong
852-34422803
[2]Dept. of Computer Science and Technology,
Harbin Institute of Technology, Shenzhen Campus,
Shenzhen, China
ruifeng.xu@cityu.edu.hk

Chunyu Kit
[1]Dept. of Chinese, Translation and Linguistics,
City University of Hong Kong
Kowloon, Hong Kong
852-27889310
ctckit@cityu.edu.hk

ABSTRACT
This paper presents the design and implementation of an opinion mining system developed by NLPCity group for NTCIR-8 MOAT evaluation, named CTL-OM. CTL-OM incorporates two opinion mining approach, namely feature-based approach and similarity-based approach. The feature-based approach incorporates computational features at punctuation-, word-, collocation-, phrase-, sentence-, paragraph- and document-level in a coarse-fine multi-pass classification framework. The opinion holders and opinions targets in the opinionated sentences are then recognized. The similarity-based approach works in a different way. This approach estimates the similarity between the example sentences and testing sentence and identifies the similar example sentence-testing sentence pair. The opinion components annotated in the example sentence are utilized to recognize the corresponding components in the testing sentence. The analysis outputs by these two approaches are integrated to obtain the final opinion mining results. CTL-OM achieved promising results in Traditional Chinese and Simplified Chinese evaluation in MOAT-8, respectively. This result shows that the incorporation of feature-based and similarity-based opinion mining approach is effective.

Categories and Subject Descriptors
I.2.7 [**Natural Language Processing**]: Text analysis

General Terms
Algorithms, Experimentation, Languages

Keywords
Opinion mining, feature-based opinion mining, similarity-based opinion mining

1. INTRODUCTION
Opinion analysis technique, which identifies and analyzes the subjective opinions and sentiments in text, attracts increasingly research interest in recent years. This technique supports many applications such as product review analysis and social news monitoring. Furthermore, this technique promotes many natural language processing tasks such as summarization and information retrieval [1,10].

Various opinion analysis techniques were investigated. Most of them can be camped into three major types. The first one, knowledge-based type, mainly utilizes the opinion-related linguistic knowledge to analyze the opinions [3]. The feature-based one is based on computational features and classifiers. Machine learning algorithms, including both supervised learning and unsupervised/semi-supervised learning algorithms, are widely adopted to improve the classifier [5]. The third one employs semantic role labeling based on FrameNet in opinion analysis [2].

However, these techniques are designed for different purposes, and their performances are evaluated based on different languages and different text set. Thus, the evaluation of opinion analysis technique has become a barrier to the research in this topic. The NTCIR Multilingual Opinion Analysis Task (MOAT) series and former Opinion Analysis Pilot Task provide an common evaluation framework on the same datasets (including English, Japanese, Simplified Chinese and Traditional Chinese text) [6,7,8]. In NTCIR-8 MOAT, there are five basic sub-tasks, including identify the opinionated sentences, identify the topic-relevant sentences, determine the polarity of opinionated sentences and recognize the opinion holder and opinion targets.

This paper describes the system (ID: CTL) developed by the NLPCity group in Dept. of Chinese, Translation and Linguistics, City University of Hong Kong in NTCIR-8 MOAT. We participated four of five subtasks, namely opinionated sentence identification, opinion polarity determination, opinion holder identification and opinion target recognition on both Traditional Chinese and Simplified Chinese dataset. Our system, named CTL-OM, incorporates feature-based and similarity-based opinion analysis techniques. In which, the feature-based approach separates the whole opinion analysis process to two stages, i.e. classification-based opinionated sentence identification and polarity determination, and extraction-based opinion holder and opinion target recognition. Firstly, computational features at punctuation-, word-, collocation-, phrase-, sentence-, paragraph- and document-level are incorporated in a coarse-fine multi-pass classification-based stage. The opinion holders and opinions targets are then analyzed in the extraction-based stage. The similarity-based approach works in a different way. This approach estimates the similarity between the example sentences and testing sentence. If a similar example-testing sentence pair is found, the opinion analysis of the testing sentence is conducted with the reference of annotated examples. It is in one stage. The

analysis results by two approaches are integrated to obtain the final outputs. CTL-OM achieved promising results in Traditional Chinese (ranked 2nd, 1st, 1st and 1st in four subtasks, respectively) and Simplified Chinese (ranked 4th, 1st, 2nd and 1st, respectively) evaluation, respectively. This result shows the incorporation of feature-based and similarity-based techniques is effective to enhance the opinion analysis system.

The rest of this paper is organized as follows. Section 2 analyzes the components of opinion expression. Section 3 presents the framework of CTL-OM. Section 4 presents the feature-based opinion analysis techniques and Section 5 presents the similarity-based techniques. Section 6 evaluates CTL-OM and finally, Section 7 concludes this paper.

2. ANALYSIS OF OPINION EXPRESSION

Opinions are usually expressed in some conventional patterns. Generally speaking, an opinion expression consists of five major components: (1) **opinion holder**, which is the governor of an opinion and normally refers to a person, a state or an organization as well as the corresponding pronouns; (2) **opinion object**, which is the target of the opinion. Most opinion objects are nouns, noun phrases or pronouns; (3) **opinion word**, which reflects the opinion polarity, i.e. positive, neutral or negative; (4) **opinion operator**, which is the verb indicating an opinion event; and (5) **opinion indicator**, which is the word indicating the orientation of an opinion or the orientation tendency of multiple opinions. In the real txt, no all of the above components always appear in each opinionated sentence.

The opinion operators are the verbs which indicating opinion expression. The typical opinion operators including 警告(warning) 强调(emphasize) and 指出 (point out). It is noted that some operators bring opinions such as 称赞(*praise*).

Opinion indicators are mainly conjunctions, adverbs and adverbial phrases, including: (1) negation conjunctions, such as 但是(*but, however*) and 尽管(*Though*), indicate that the sentiment of the following clause/sentence is different from the preceding one; (2) continual conjunctions, such as 并且, 而且(and) and 特别 (*especially*), indicate that the sentiment of the following clause/sentence is the same as the preceding one; (3) adverbs and adverbial phrases directly indicate the polarity of the opinionated sentence, e.g. 遗憾的是 (*It is regrettable*); and (4) verbs directly indicate the polarity of the opinionated sentence.

Opinion words play a key role in opinion expression. They are generally classified into: (1) c**ontext-free opinion word** (***CFOW***) whose polarity is constant irrespective of context, e.g. 完美 (*perfect*) is absolute positive and 恶劣(*bad*) is negative; (2) **context-dependent opinion word** (***CDOW***) whose polarity is determined by their context. For example, 好笑的 is positive when it is used in the context of talk shows (meaning *burlesque*); but it is negative when it is used in the context of politics (meaning *absurd)*; and (3) **object-dependent opinion word** (***ODOW***) is the neutral word carrying different polarities when associated with different opinion objects. For example, 高(*high*) expresses positive sense when collocating with 性能(*performance*) but brings negative sense when collocating with 债务(*debt*). For practical reasons, this kind of words are always processed as CDOWs.

3. FRAMEWORK DESIGN

CTL-OM separates the four opinion analysis subtasks into two groups: the classification-based opinionated sentence identification and polarity determination, and the extraction-based opinion holder and opinion target recognition. CTL-OM consists of two components. The first one is feature-based, which is an improved version of our previous system adopted in NTCIR-7 MOAT [9]. Computational features at different levels are extracted from testing sentences. These features are then incorporated in a coarse-fine multi-pass classification system to identify the opinionated sentence and determine their polarity. Next, based on these features and syntax and rules, the opinion holders and targets are recognized. The second component in CTL-OM is based on similarity between example sentence and testing sentence. If a similar example sentence corresponding to the testing sentence is found, the testing sentence analysis is conducted based on the opinion component matching provided in the annotated example sentence. This approach is obviously different from the first one. The whole analysis process regards a sentence as a whole of sequent words rather than separated features in the feature-based opinion mining approach. The analysis results of these two components are integrated to generate the final output. The whole framework is illustrated in Figure 1.

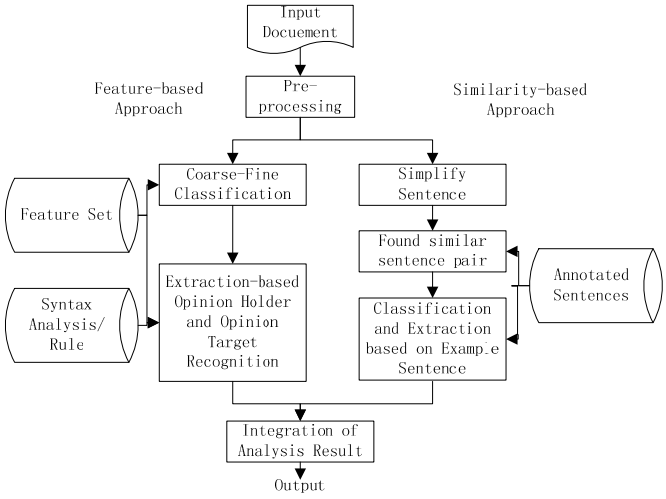

Figure 1. Framework of CTL-OM.

4. FEATURE-BASED OPINION ANALYSIS

4.1 Feature Selection

Based on the observation and statistic analysis, a set of computational features related to opinion analysis at different levels are designed and selected.

Punctuation level features:

The presence of direct quote punctuation “「” and “」”

Word-level and entity-level features:

The presence of known opinion operators: E.g. 银行**称**此项政策可以缓解资金紧张问题 (*Bank **said** this policy can ease the financial constraint*)

The presence of known context-free opinion words/the percentage of known context-free opinion words in sentence: E.g. 演出非常**完美** (*The performance is **perfect***)

The percentage of known opinion words in sentence/the percentage of known opinion word in sentence: E.g. 她说, 中国在此次泰国 DNA 检测工作中发挥了**重要**作用 (*She said that, China plays an **important** role in the DNA testing work in Thailand*)

The presence of named entity/the percentage of named entity in the sentence: E.g. 她说, **中国**在此次**泰国** DNA 检测工作中发挥了重要作用 (*She said that, **China** plays an important role in the DNA testing work in **Thailand***)

The presence of pronoun/the percentage of pronoun in the sentence: E.g. **她**说,中国在此次泰国 DNA 检测工作中发挥了重要作用 (***She** said that, China plays an important role in the DNA testing work in Thailand*)

The presence of known opinion indicators: E.g. 他**称赞**这项技术可以带来巨大效益 (*He **praised** this technique may bringing tremendous benefits*)

The presence of known degree adverbs: E.g. 他称赞这项技术可以带来**巨大**效益 (*He praised this technique may bringing **tremendous** benefits*)

Collocation-level features:

The presence of collocations between named entities and opinion operators: E.g. **银行 称** 此项政策可以缓解资金紧张问题 (***Bank said** this policy can ease the financial constraint*)

The presence of collocations between pronouns and opinion operators: E.g. **她说**,中国在此次泰国 DNA 检测工作中发挥了重要作用 (***She said** that, China plays an important role in the DNA testing work in Thailand*)

The presence of collocations between nouns or named entities and opinion words: **演出**非常**完美** (***The performance** is **perfect***)

The presence of collocations between pronouns and opinion words: E.g. **他**很**优秀** (***He** is very **good***)

The presence of collocations between degree adverbs and opinion words: E.g. 他**很优秀** (*He is **very good***)

The presence of collocations between degree adverbs and opinion operators: E.g. 中国**强烈谴责**这一罪行 (*China **strongly condemns** this crime*)

Sentence-level features:

The opinions of neighboring -2, -1, +1, and +2 sentences:

The transition probabilities from opinionated (or factual)/polarity of neighboring sentences to possible states of current sentences (opinionated or not, polarity):

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,

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

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. The conditional probabilities obtained from training corpus are used as transition probabilities in this stage.

Paragraph-level features:

The opinionated/fractural probability of current paragraph:

The polarity of current paragraph:

Document-level features:

The opinionated/fractural probability of current document:

The polarity of current document:

4.2 PRE-PROCESSING

Word segmentation and Part-of-Speech tagging are indispensable steps in Chinese sentence analysis. A Unicode-based word segmentation and POS tagging system proposed in [4] are adopted. It is trained by 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 entities are recognized. The identified named entities are helpful to recognize the opinion holders and opinion targets.

Another task in pre-processing is to recall several pre-defined types of omitted components in continuous sentences. Here, we give an example.

Example Type 1: Multi-Sentences in Quotation

Sentence 1.反对党则指出，“苏哈托家族贪污情况非常严重。(*Opposition party pointed out that " Suharto family's corruption is very serious.*)

Sentence 2. 此外，苏哈托在其统治期间涉嫌多宗人权损害事件。”(*In addition, Suharto suspected some cases of human rights affection during his govern*"

The holder of these two sentences are both 反对党(*Opposition party*) while the holders does not appear in the second sentence. This holder will be appended in the second sentences during preprocessing stage. Such pre-processing is helpful to improve the opinion holder recognition.

4.3 Coarse-Fine Classification

Most existing opinion analysis techniques regard opinionated sentence identification as one-step classification problem. The linguistic features and computational features in the testing sentence are utilized in the classifier to determine its opinion and

polarity. These techniques ignore the clues of opinions of the paragraph, document and the neighboring sentences to the opinion of current sentence. The observation on NTCIR-6 and NTCIR-7 opinion corpus shows that 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. Naturally, the paragraph-, document-level and neighboring sentence-level opinions should be considered in the sentence opinion analysis. Meanwhile, humans normally analyze 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 classification framework. This framework has multi-pass coarse-fine analysis which is described below.

Input: Document D consists of sentences $S_0, S_1, S_i \ldots S_n$

Step 1. Use the base classifier, C_{base}, to analyze the opinion of each sentence in D. The output is the polarity value, $Pol(S_i)$. In this step, a linear classifier is adopted. Meanwhile, the features at punctuation-level, word/named entity level and collocation-level are incorporated. Noted here that, all of the features adopted in this step are inner-sentence ones.

Step 2. Estimate the polarity of D and each paragraph P.

$$Pol(D) = \frac{1}{n} \cdot \sum_{i=1}^{n} Plo(S_i) \tag{2}$$

$$Pol(P) = \frac{1}{j} \cdot \sum_{i=1}^{j} Plo(S_i) \tag{3}$$

Step 3. Use the improved classifier, C_{im}, to estimate the opinion of each sentence, $Pol(S_i)$*. C_{im} incorporates paragraph-, document- and neighboring sentence-level opinions as new features. Here, the transition probabilities are obtained from training data.

Step 4. Update the document and paragraph opinion using $Pol(S_i)$*.

Step 5. If the sentence and document opinion mining output converge, terminate. Otherwise, go to Step 3.

The output of this stage is the identified opinionated sentences and their polarity.

4.4 Extraction-based Opinion Holder and Target Recognition

Firstly, 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, dependency, conjunctions, suffix, prefix and opinion operator, respectively. Furthermore, some syntax rules based on preposition phrase (related to opinion holder identification) are adopted. Finally, the position of the opinion holder candidate in the sentence and the respective position to the opinion operator candidate are considered as well as their dependency.

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 22 rules and patterns for opinion targets of persons and 35 patterns for recognizing clause-level opinion targets.

5. SIMILARITY-BASED OPINION ANALYSIS

Based on the observation on the opinion corpus, we adopt a hypothesis that in similar type of texts, opinionated sentences will be more similar to other opinion sentences than to factual sentences [10]. Especially, we found that, in two similar opinionated sentences, their opinion holders and opinion targets always have the same or similar places. These observations motivate the development of similarity-based opinion analysis technique. That is, for an input sentence, looking for a similar example sentences in the annotated opinion corpus. The opinion determination and opinion holder/opinion target recognition may be done by matching the corresponding components in example sentence to testing sentence.

5.1 Sentence Simplification

Generally speaking, the sentence similarity can be measured based on shared words and phrases among two sentences. However, there is a big barrier to apply similarity-based approach to opinion analysis that the size of annotated opinion corpus is not large enough. The sparseness problem affects the coverage of similarity-based approach. Therefore, we must simplify both the testing sentences and examples sentences.

Firstly, we simplify some words by assign a synonyms set id to substitute a group of synonyms. Such operation utilizes the knowledge in TongYiCi CiLin (A Chinese synonyms dictionary). For example, 烦恼(*trouble*) have following synonyms 烦闷、烦心、糟心、烦乱、烦躁. They are assigned id Syn147. The following two sentences are then simplified as,

在我们这个年代，有许多**烦恼** (*There are a lot of troubles in this age*) ->在我们这个年代，有许多 **Syn147**

这场火灾使她十分**烦乱** (*The fire quite upset her*) ->这场火灾使她十分 **Syn147**

Secondly, we substitute frequent opinion operators, degree adverbs and context-free opinion words (positive, negative) by the labels of *oper*, *dead*, *cfow(-p,-n)*, respectively. For example, 他很优秀 (*He is very good*) –> 他 *dead cfow-p.*

Thirdly, we substitute several types of noun phrases, including organization names, country names, place name and person names, etc. by the corresponding labels. For example, 中国强烈谴责这一罪行(*China strongly condemns this crime*) –> country_name 强烈谴责这一罪行.

Finally, we apply a chunker to identify and substitute base quantifier phrases (e.g. 数千名 *thousands of*->bqp), base time phrases (e.g. 早上 8 时 *8:00 in the morning*->btp), base position phrases (e.g. 内蒙古东北部 *North-east of Inner Mongolia*->bpp) in the sentences.

In this way, many sentences are simplified. Such an operation is expected to enhance the coverage of common components during sentence similarity estimation.

5.2 Sentence Similarity Estimation

We applied two methods to estimate the sentence similarity. Firstly, we estimate the coverage of common words/simplified labels between testing sentence and example sentence. Supposes that sentences *a and b* have n_a and n_b words/simplified labels, respectively, in which m ones are common, their similarity, $sim(a,b)$, is estimated by,

$$\mathrm{Sim}(a,\ b) = \frac{m}{n_a + n_b} \tag{4}$$

Secondly, we estimate the edit distance between two sentences, which is the lowest operation cost for varying sentence *a* to sentence *b* through *substitution*, *insertion* and *deletion* operations. Currently, we set the cost for each of above three operations as equal.

The two similarities are incorporated to generate final sentence similarity. The sentences pairs having similarity greater than a empirical threshold are retrieved.

5.3 Opinion Analysis based on Similar Sentences

Once a similar example sentence corresponding to testing sentence is found, the testing sentence is analyzed by using the information annotated on the example sentence. If the similar example sentence is an opinionated one, the same polarity will be assigned to the testing sentence. Meanwhile, the words with same part-of-speech/simplified labels at similar positions in the example-testing sentence pair are matched. The components in the testing sentences corresponding to the annotated opinion holder and opinion target in the example sentence are recognized as the opinion holder and opinion target of the testing sentence.

5.4 Combine Feature-based and Similarity-based Analysis Results

Considering that the similarity-based analysis approach has the advantages on opinion holder and opinion target recognition but relative low performance on opinionated sentence identification, we adopt the opinion classification results of sentences in the final output. The recognized opinion holder and opinion targets by feature-based and similarity-based approach are combined to generate final output, in which emphasizes the similarity-based analysis results.

6. EVALUATIONS AND DISCUSSIONS

CTL participate the evaluation on Traditional Chinese (TC) side and Simplified Chinese (SC) side at NTCIR-8 MOAT. The descriptions on the training data, testing data and gold standard are given in the overview paper of NTCIR-8 MOAT [8]. We participate four subtasks, including, opinionated sentence identification, opinion polarity determination, opinion holder and opinion target recognition. The organizer evaluates the performance of the systems by using precision (P), recall (R) and F-measure (F).

First, the opinionated sentence identification subtask is evaluated. The performances achieved by CTL-OM on TC and SC side are given in Table 1, respectively, as well as their ranks. For reference, the performances of the system with highest F are listed.

Table 1. The evaluation on opinionated sentence identification

	P	R	F	Rank
TC				
CTL	65.14	68.79	66.92	2nd
Best	56.39	85.71	68.05	1st
SC				
CTL	36.46	78.90	49.87	4th
Best	41.34	83.35	55.27	1st

It is shown that the CTL-OM's performance on this subtask is not satisfactory, especially on the SC side. CTL achieves highest precision on TC side, i.e. 0.6514, which is obviously higher than the second-highest precision, 0.5639, achieved by CityUHK group. However, the recall is much lower. It shows that our system should be revised to make a trade-off between precision and recall. As for the SC side, the achieved precision is obviously lower than TC side. It partially attributes to the annotation disagreement on the gold answer between SC and TC. Especially, we found that the percentage of annotated opinionated sentence on TC side is nearly the double of SC side. CTL-OM is Unicode-based which analyzes SC and TC text in a same system and combined training data. Thus, the achieved performances on TC and SC sides are obviously.

Second, the polarity determination is evaluated. The performances achieved by CTL-OM and by reference system are given in Table 2, respectively.

It is shown CTL achieves both the highest precision and the best F both on TC and SC side. This result benefits from the coarse-fine opinion classification which incorporates both inner- and inter-sentence features and coarse-fine multi-pass classification strategy.

Table 2. The evaluation on polarity determination

	P	R	F	Rank
TC				
CTL	76.50	53.06	62.66	1st
Best	76.50	53.06	62.66	1st
SC				
CTL	67.39	52.90	59.27	1st
Best	67.39	52.90	59.27	1st

Finally, the performance on opinion holder and opinion target recognition are evaluated. The performances achieved by CTL-OM and by reference system are given in Table 3 and Table 4, respectively.

Table 3. The evaluation on opinion holder recognition

	P	R	F	Rank
TC				
CTL	84.90	-	-	1st
Best	84.90	-	-	1st
SC				
CTL	95.30	73.20	82.80	1st
Best	87.70	79.20	83.20	2nd

Table 4. The evaluation on opinion target recognition

	P	R	F	Rank
TC				
CTL	54.40	-	-	1st
Best	54.40	-	-	1st
SC				
CTL	73.50	56.40	63.80	1st
Best	73.50	56.40	63.80	1st

It shows that CTL-OM achieves encouraging performance on opinion holder and opinion target recognition. The results indicate that incorporation of feature-based and similarity-based techniques are effective.

7. CONCLUSIONS

In this paper, we present the CTL-OM system developed for NTCIR-8 MOAT task. This system incorporates feature-based coarse-fine classification and analysis approach, which separate opinionated sentence identification/polarity determination and opinion holder /opinion target in two stages, and similarity–based approach, which performs four subtasks in one stage, in opinion mining. The analysis results by different approach are combined to enhance the performance of the final output. The evaluations on Traditional Chinese side and Simplified Chinese side in NTCIR-8 MOAT show the effectiveness of the proposed opinion analysis framework, especially on the opinion holder/opinion target recognition. The future researches will focus on the improvement on the two individual analysis approaches and better integration strategy.

8. ACKNOWLEDGMENTS

This paper is supported by a post-doctoral research fellowship funded by City University of Hong Kong, Hong Kong.

9. REFERENCES

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

[2] Kim, S.M. and Hovy, E. 2006. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text, In Proc. ACL/COLING Workshop on Sentiment and Subjectivity in Text, pp. 1-8, Sydney, Australia

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

[4] Lu, Q., Chan, S. T., Xu, R. F. et al. 2003. A Unicode based Adaptive Segmentor. In Proceedings of the 2nd SIGHAN Workshop on Chinese Language Processing at ACL 2003, pp. 164-167, Spain

[5] Pang, B., Lee, L. and Vaithyanathan S. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proc. EMNLP 2002*, pp.79–86.

[6] Seki, Y., Evans D.K., et al. 2007. Overview of Opinion Analysis Pilot Task at NTCIR-6. In Proc. of NTCIR-6 Workshop. Japan.

[7] Seki, Y., Evans D.K., et al. 2008. Overview of Multilingual Opinion Analysis Task at NTCIR-7. In Proc. of NTCIR-7 Workshop. Japan.

[8] Seki, Y., Ku, L.W., et al. 2010. Overview of Multilingual Opinion Analysis Task at NTCIR-8. In Proc. of NTCIR-8 Workshop. Japan.

[9] Xu, R.F., Wong, K.F. and Xia, Y.Q. 2008. Coarse-Fine Opinion Mining – WIA in NTCIR-7 MOAT Task, In Proc. of NTCIR-7 Workshop. Japan.

[10] Yu, H. and Hatzivassiloglou, V. 2003. Towards Answering Opinion Question: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences, In EMNLP 2003