Computational Linguistics and Chinese Language Processing
Vol. 15, No. 1, March 2010, pp. 37-60



# Tourism-Related Opinion Detection and Tourist-Attraction Target Identification

**Chuan-Jie Lin[∗] and Pin-Hsien Chao[∗]**

## Abstract

This paper focuses on tourism-related opinion mining, including tourism-related opinion detection and tourist-attraction target identification. The experimental data are blog articles labeled as being in the domestic tourism category in a blogspace. Annotators were asked to annotate the opinion polarity and the opinion target for every sentence. Different strategies and features have been proposed to identify opinion targets, including tourist attraction keywords, coreferential expressions, tourism-related opinion words, and a 2-level classifier. We used machine learning methods to train classifiers for tourism-related opinion mining. A retraining mechanism is proposed to obtain the system decisions of preceding sentences. The precision and recall scores of tourism-related opinion detection were 55.98% and 59.30%, respectively, and the scores of tourist attraction target identification among known tourism-related opinionated sentences were 90.06% and 89.91%, respectively. The overall precision and recall scores were 51.30% and 54.21%, respectively.

**Keywords:** Tourism-Related Opinion Mining, Tourist Attraction Target Identification, Opinion Analysis.

## 1. Introduction

The blogspace is a large resource for opinion mining. Opinion extraction methods are valuable for a wide range of applications.

Our initial interest is to extract opinions related to tourist attractions from blog articles because it is helpful to see other people's opinions about tourist attractions when planning a tour. Nevertheless, two issues arise when trying to apply published methods to retrieve opinions of tourist attractions:

[∗] Department of Computer Science and Engineering, National Taiwan Ocean University
No 2, Pei-Ning Road, Keelung, 20224 Taiwan
E-mail: {cjlin, M96570038}@ntou.edu.tw

(1) Sentence-level or document-level:

A travel article is often multi-topic because a travel route often includes several tourist attractions. Therefore, the opinion analysis for a specific tourist attraction should be carried out at sentence level, not document level.

(2) Opinion topic or opinion target:

Tourist attractions may be treated as topics (queries in IR) or as targets of opinions. Consider the following two sentences selected and adapted from our dataset:

*The Dream Lake is a beautiful place.*
*The water is green and clear.*

Both sentences are considered tourism-related opinions by us. Their opinion targets, however, are not the same. The opinion target of the first sentence is “the Dream Lake” itself, while the target of the second sentence is “the water” (in the Dream Lake). Both sentences are related to the same topic, “the Dream Lake,” but the second sentence does not contain its topic words. We find difficulty in applying the previously developed methods due to these reasons.

Opinion mining and analysis have been widely studied in several topics, including opinion detection and polarity classification (Wiebe *et al*., 2001; Pang *et al*., 2002; Alm *et al*., 2005; Ghose *et al*., 2007), opinion holder finding (Choi *et al*., 2005; Kim & Hovy, 2005; Breck *et al*., 2007), and opinion summarization (Ku *et al*., 2005). Some well-known large-scale opinion mining benchmarks have also been created, such as the NTCIR MOAT datasets (Seki *et al*., 2010) which are constructed on four languages, including Traditional Chinese.

Opinion retrieval is one of the research topics relevant to our work. Godbole *et al.* (2007) estimated the polarity scores for a large set of named entities. Nevertheless, the opinionated sentences that did not contain named entities were skipped because they measured the scores by the co-occurrences of named entities and opinion words. Ku *et al.* (2005) retrieved documents containing relevant opinions relating to TREC-like topics. Zhang *et al.* (2008) accepted short queries (titles only) and expanded the queries by web resources and relevance feedback. The units of their retrieval work, however, were documents, not sentence-level. Okamoto *et al.* (2009) extracted relevant opinionated sentences by language model. Unfortunately, a large-scale training set is required to build a reliable probabilistic model, which is labor-consuming to prepare in the tourism domain.

Opinion target identification is another research topic that is relevant to our work. Many researchers have focused on learning features of pre-defined types of products from reviews (Hu & Liu, 2004; Ghani *et al*., 2006; Xia *et al*., 2009). Nevertheless, the question remains whether the features of all kinds of tourist attractions are common. Moreover, in the conventional definition, an opinion target in a tourism-related opinion is not always the name of the tourist attraction.

Therefore, we define tourism-related opinion mining as a new topic and propose several approaches to solve the problem, including rule-based approaches and machine learning approaches. Although the experimental data used in this paper are written in Chinese, many of the rules and features are not language-dependent or can be easily adopted if necessary resources are available. We also hope that the experience gained from these experiments can be applied to other domains where articles are often multi-topic, such as baseball game critics.

The structure of this paper is as follows. Section 2 presents the main ideas of tourism-related opinion identification and introduces the resources prepared for the work. Section 3 describes the design of a rule-based opinion identification system. Section 4 defines the features for training classifiers to build an opinion identification system. Section 5 discusses the experimental results, and Section 6 concludes this paper.

## 2. Tourism-Related Opinion Analysis

### 2.1 Problem Definition

Opinionated sentences related to tourist attractions are the main interest of this paper. We call such an opinionated sentence a ***tourism-related opinion*** (hereafter "***TR-opinion***") and its targeted tourist attraction a ***tourist attraction target*** (hereafter "***TA-target***").

The main goal of this paper is to retrieve TR-opinions and determine their TA-targets. That is, given an opinionated sentence, determine whether it is tourism-related or not, and decide which tourist attraction is the focus of this opinion. Our experiments were performed based on two assumptions: (1) sentences have been *correctly* tagged as 'opinionated' or not; (2) tourist attraction names appearing in a document have been *correctly* recognized. Hence, we have not integrated an opinion detection module and a tourist-attraction recognition module into our system yet.

Opinion identification is not the main focus of this paper. There has been a lot of research on this topic. In the future, we would like to perform well-developed methods to do opinion detection in order to build a full system. In this paper, though, the input sentences are those sentences correctly labeled as opinions.

Tourist attraction name recognition also is not a focus of this paper. It requires a named entity recognition system specifically designed for tourist attraction names, but we cannot find one. Although some of the tourist attractions are locations or organizations, such as parks or museums, there are various types of names, such as monuments or scenic spots that would need to be learned. In this paper, we simply prepare a list of tourist attraction names and manually check the correctness of the occurrences of the attraction names in the articles.

Tourist attraction name recognition will be studied in the future.

The main ideas in accomplishing the tasks are:

(1) Some opinion words strongly hint that a sentence is tourism-related.

(2) The frequency of use of a tourist attraction and its distance to an opinionated sentence can be useful information.

(3) A tourist attraction can be expressed in several ways in an article. This is the well-known coreference problem.

(4) A sentence may target a tourist attraction if its preceding sentence also focuses on a tourist attraction.

Before designing rules or features according to these ideas, some resources were prepared beforehand, as described in the following subsections.

## 2.2 Experimental Dataset Preparation

The best known benchmarks for opinion mining are the NTCIR MOAT datasets (Seki *et al.*, 2010). There was one pilot task in NTCIR-6 and were two formal tasks in NTCIR-7 and NTCIR-8. There are a total of 70 topics in Traditional Chinese. Nevertheless, none of their information need is about tourism attraction opinions. Although some topics may bring in tourism-related documents, such as the terrorist bombing on Bali Island and the tsunami in Sumatra, the number of topics is too small, and we still have to find TR-opinions among the opinionated sentences. For these reasons, we decided to build a new experimental dataset in the tourism domain.

200 travel articles were collected from a blog site called Wretch[1] (無名小站). These articles were categorized as “domestic travel” on the blog site. We chose the most recommended articles by the readers in order to assure that the articles were truly about travel.

Three annotators were asked to annotate the data. Each sentence was labeled as opinionated or not, its opinion polarity was assigned, and its TA-target was found if the annotator considered it a TR-opinion.

The guidelines of TA-target decision for the annotators are as follows. Given a document, a list of tourist attractions mentioned in the document is shown to the annotators. A TA-target must be one of the tourist attractions on the list. If an opinion is made on a part of a tourist attraction (*e.g*. the souvenir shop in an amusement park), its TA-target is set to be the tourist attraction. If an opinionated sentence mentions a tourist attraction together with the city it belongs to, its TA-target is set to be the tourist attraction only. A city can be chosen as a TA-target only when the blogger directly expresses his or her feeling about the city. Note that,

[1] http://www.wretch.cc/blog

if a sentence only expresses the blogger's emotion (*e.g*. "*I am so happy today*"), it is not a TR-opinion.

The final annotations of the experimental dataset were determined by two-stage voting. The first stage determined a sentence being positive-, neutral-, negative-, or non-opinionated. The second stage determined the sentence being a TR-opinion or not by deciding its TA-target. In each stage, an option agreed upon by at least two annotators became the final annotation. If no agreement was found, the authors of this paper would choose one of the decisions made by the annotators. Those sentences voted as "non-opinionated" in the first stage were automatically labeled as "not TR-opinion" in the second stage.

***Table 1. Agreements of Data Annotations***

| **Comparison** | **Opinion and Polarity** | **TR-opinion** | **TA-target** |
|---|---|---|---|
| Annotator 1 vs. 2 | 0.608 | 0.569 | 0.568 |
| Annotator 1 vs. 3 | 0.584 | 0.518 | 0.518 |
| Annotator 2 vs. 3 | 0.589 | 0.529 | 0.529 |
| Exp Data vs. A1 | 0.791 | 0.761 | 0.761 |
| Exp Data vs. A2 | 0.792 | 0.769 | 0.769 |
| Exp Data vs. A3 | 0.758 | 0.701 | 0.701 |

Table 1 lists the agreement of TR-opinion and TA-target measured by Cohen's kappa coefficient. The first three rows show the agreement among the annotators. The last three rows give the agreement between the final experimental dataset and each annotator. We can see that the agreement level is not high enough. This means TR-opinion detection and TA-target identification are very challenging.

Among the 200 articles, 37 of them did not contain a tourist attraction and 7 did not contain a TR-opinion. After removing these articles, there were a total of 10,904 sentences in the remaining 156 articles, with 3,542 opinionated sentences and 1,199 TR-opinions, which leads to a precision rate of 33.9% (1199/3542) if a baseline system guesses all of the opinions as TR-opinions.

Table 2 lists the statistical data regarding the number of tourist attractions mentioned in the articles. As we can see, 28 articles contained only one tourist attraction, which means that almost 89% of the articles mentioned multiple tourist attractions, making TA-target detection an issue. There were on average 6.378 tourist attractions mentioned in each article.

***Table 2. Number of Tourist Attractions in Articles***

| **#TA** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11~20 | 21~78 | Average |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **#docs** | 28 | 19 | 23 | 12 | 13 | 14 | 9 | 5 | 6 | 3 | 17 | 7 | 6.378 |

## 2.3 Tourism-Related Opinion Words

Some opinion words are more related to tourist attractions than others. Consider the following two examples:

*I am so excited that the vacation is coming.*
*The lake is so large and clear.*

The adjective "excited" is often used when describing personal feelings. On the other hand, "clear" is often seen in sentences describing scenic spots. We can say that opinion words are often domain-dependent.

Many papers have focused on finding domain-specific opinion words and deciding their polarities, as mentioned in Section 1. This, however, is slightly different from our need. "Domain" in their works often refers to "a product type," such as *digital cameras*. Opinion words related to digital cameras are the adjectives used to express the features of digital cameras, such as "long" for *battery life* and "heavy" for *weight*.

Nevertheless, the question remains as to whether there are common features or attributes among tourist attractions. The feature *water* or *clearness* only relates to bodies of water, such as rivers and lakes, while the feature *design* only relates to buildings. Moreover, there are many adjectives expressing opinions directly without denoting any specific features, such as *amazing* and *beautiful* (*e.g*. "this city is beautiful"). Therefore, we want to collect a set of opinion words which are often used in tourism-related opinionated sentences without considering features.

We define a simple function ***TRscore*(*ow*)**, the *tourism-relatedness score*, to estimate the likelihood of an opinion word *ow* appearing in a TR-opinion by evaluating the ratio of the opinionated sentences where the word *ow* appears to be tourism-related:

$$TRscore(ow) = \frac{\#(ow \text{ in TR - opinion})}{\#(ow \text{ in opinion})} \tag{1}$$

Opinion words whose TR-scores are higher than a predetermined threshold are collected as the ***tourism-related opinion words*** (hereafter "***TR-opword***"). The determination of the value of the threshold of TR-scores is discussed in Section 5.1.

## 2.4 Coreferential Expressions

Coreference is an important problem in natural language processing. When a tourist attraction is mentioned in an article, it is quite often expressed in several different ways. Consider the following three sentences selected and adapted from our experimental dataset:

> *My family and I visited the Wufeng Resort last week.*
> *We were impressed by the fresh air when we arrived at the resort.*
> *Wufeng also thoughtfully provides parking service.*

All three underlined expressions refer to the same tourist attraction "*the Wufeng Resort*," where "*resort*" is its category, "*Wufeng*" its name, and "*the Wufeng Resort*" its full name.

It is quite common to refer a tourist attraction by the category keyword in its name. For this reason, we created a list of ***tourist attraction keywords*** (hereafter ***TA-keywords***), which are tourist attraction categories. Note that there are several synonymous keywords in the same category. The method of collecting TA-keywords is as follows.

First, a tourism website called Travel King[2] (旅遊資訊王) was visited and 1,836 tourist attraction names located in Taiwan were collected. All of the names were written in Chinese without word segmentation.

For every pair of tourist attraction names, their longest common trailing substring was extracted. The substrings containing only one Chinese character were discarded. After having humans check their correctness, 158 TA-keywords were collected, such as 國家公園 (national park) and 溫泉 (hot spring).

We do not resolve the coreference problem directly. Instead, we try to find potential coreferential expressions. The frequency or distance feature of a tourist attraction is measured by the occurrences of all kinds of coreferential expressions of this tourist attraction. The first type of coreference is expressed by the longest TA-keyword found in a tourist attraction's name.

The list of the TA-keywords may not be complete enough. Some types of names are not in the list. In order to make the system more robust, we also take the trailing substring (the last two characters) of a full name as one of its possible coreferential expressions.

Similarly, although we can extract the name part of a tourist attraction by deleting the keyword part from its full name, we simply take its leading substring (the first two characters) as one of its possible coreferential expressions.

The function $ref_{all}$(a) is defined to denote all possible coreferential expressions of a tourist attraction *a*. For example, $ref_{all}$(五峰渡假村) = {五峰渡假村, 渡假村, 五峰, 假村}, *i.e*. for the tourist attraction 五峰渡假村, its possible coreferential expressions include its full name "五峰渡假村" (*the Wufeng Resort*), its TA-keyword "渡假村" (*Resort*), its leading substring "五峰" (*Wufeng*), and its trailing substring "假村". An example of coreferential expression detection is given here:

[2] http://travel.network.com.tw/tourguide/twnmap/

我和家人上星期去$\boxed{\text{五峰渡假村}}$玩

(My family and I visited $\boxed{\text{the Wufeng Resort}}$ last week.)

一到<u>渡假村</u>$_1$就對那邊的新鮮空氣印象深刻

(We were impressed by the fresh air when we arrived at <u>the resort</u>$_1$.)

<u>五峰</u>$_2$也貼心地提供了停車的服務

(<u>Wufeng</u>$_2$ also thoughtfully provides parking service.)

如果只是單純的放鬆自己什麼都不想

(If you simply want to relax and get away from it all,)

$\boxed{\text{五峰渡假村}}$是個不錯的選擇

($\boxed{\text{the Wufeng Resort}}$ will be a good choice.)

In this paragraph, a full name "*the Wufeng Resort*" (the bordered text) appears in the first and the last lines, while its TA-keyword "*resort*" (the first underlined text) is found in the second line and its leading substring "*Wufeng*" (the second underlined text) in the third line.

The strategy for finding occurrences of tourist attractions in a sentence is longest-expression-first. In other words, given a set of tourist attractions $\{A_1, A_2, \ldots, A_m\}$, we will find the attraction $A_i$ whose coreferential expression appearing in this sentence is the longest.

This strategy has its limitations. If a tourist attraction does not reveal its category in its name, it would be difficult to know its category, such as *the Louvre* as a museum. Another limitation is to know the hierarchy of the tourist attractions. For example, some people will refer to the Wufeng Resort as a *hotel* or a *park*. How to detect a tourist attraction and identify its category will be our future work.

## 3. Rule-Based Approaches

To describe our approaches more clearly, Table 3 lists the definitions of notations and functions used in this paper to define opinion-mining rules and features.

The set of opinionated sentences $\boldsymbol{S_{op}}$ and the set of tourist attractions $\boldsymbol{TA}$ appearing in a document $D$ are given in advance. Our goal is to predict a set of TR-opinions $\boldsymbol{S_{to}}$ as similar to the correct set $\boldsymbol{S^{\#}_{to}}$ as possible, and determine each TR-opinion's TA-target. Note that we have $n$ sentences and $m$ tourist attractions in a document D, and $\boldsymbol{S^{\#}_{to}} \subseteq \boldsymbol{S_{op}} \subseteq \boldsymbol{S}$.

Our rule-based approaches for TR-opinion mining include the following decisions:

(1) Select a set of TR-opinion candidates $\boldsymbol{S_c}$. We can consider only a subset of the opinionated sentences $\boldsymbol{S_{op}}$ as potential TR-opinions.

(2) Select a set of TA-target candidates $\boldsymbol{TA_c}$. We can take only a subset of tourist attractions $\boldsymbol{TA}$ as TA-target candidates.

***Table 3. Notations and Functions for Defining Rules and Features***

| **Notation** | **Definition** |
|---|---|
| $\boldsymbol{S}$ | $\{S_1, S_2, ..., S_n\}$, the set of sentences in a document $D$ |
| $\boldsymbol{TA}$ | $\{A_1, A_2, ..., A_m\}$, the set of tourist attractions appearing in $D$ |
| $\boldsymbol{OW}$ | $\{ow_1, ow_2, ..., ow_p\}$, the set of known TR-opwords |
| $\boldsymbol{S_{op}}$ | the set of known opinionated sentences in $D$ |
| $\boldsymbol{S^{\#}_{to}}$ | the set of known TR-opinions in $D$ |
| $trg(s)$ | the TA-target of a TR-opinion $s$ |
| $freq(a)$ | the frequency of a tourist attraction $a$,<br>normalized by the maximal tourist attraction's frequency in $D$ |
| $\boldsymbol{A_{maxf}}$ | $\arg\max_{a \in TA} freq(a)$, the set of the most frequent tourist attractions in $D$ |
| $ref_{all}(a)$ | the set of all possible coreferential expressions of a tourist attraction $a$ |
| $in(x, j, k)$ | 1 if a string $x$ appears in one of the sentences $S_j$, $S_{j+1}$ ..., $S_k$; 0 otherwise |
| $fst(x, j, k)$ | the index of the first sentence in $S_j$, $S_{j+1}$..., $S_k$ which contains a string $x$;<br>$\infty$ if none of the sentences contains $x$ |
| $lst(x, j, k)$ | the index of the last sentence in $S_j$, $S_{j+1}$..., $S_k$ which contains a string $x$;<br>0 if none of the sentences contains $x$ |
| $Nop_{-}(S_i)$ | $\max_{k<i, S_k \in S_{op}}(k)$, the ID of the nearest opinion which precedes $S_i$;<br>-1 if no preceding opinionated sentences |
| $Nop_{+}(S_i)$ | $\min_{i<k, S_k \in S_{op}}(k)$, the ID of the nearest opinion which follows $S_i$;<br>$\infty$ if no following opinionated sentences |
| $Sid_{-}(a, S_i)$ | $\max_{x \in ref_c(a)} lst(x, 1, i-1)$, the ID of the nearest opinionated sentence which precedes $S_i$ and contains $a$ |
| $Sid_{+}(a, S_i)$ | $\min_{x \in ref_c(a)} fst(x, i+1, n)$, the ID of the nearest opinionated sentence which follows $S_i$ and contains $a$ |
| $Nid_{-}(S_i)$ | $\max_{a \in TA_c} Sid_{-}(a, S_i)$, the ID of the nearest sentence that contains a tourist attraction and precedes the sentence $S_i$ |
| $Nid_{+}(S_i)$ | $\min_{a \in TA_c} Sid_{+}(a, S_i)$, the ID of the nearest sentence that contains a tourist attraction and follows the sentence $S_i$ |

(3) Select a function of possible coreferential expressions $ref_c(a)$ of a tourist attraction $a$. We can consider only some types of expressions as coreferences to the tourist attraction $a$.

(4) Determine if a sentence $s$ in $\boldsymbol{S_c}$ is a TR-opinion.

(5) Determine which tourist attraction $a$ in $\boldsymbol{TA_c}$ is the TA-target of a TR-opinion $s$.

Two TR-opinion mining rules, *R*nt1 and *R*nt2, are proposed to guess a sentence $S_i$ in $\boldsymbol{S_c}$ being a TR-opinion and its TA-target. Their definitions are explained here as illustrated in Table 4.

**Nearest Preceding Tourist Attraction Rule (*R*nt1)**: If there is a TA-target candidate appearing inside or before $S_i$, it is predicted as a TR-opinion and its TA-target is the nearest tourist attraction.

**Nearest in-Window Tourist Attraction Rule (Rnt2)**: Set the window size as b sentences. If there is a TA-target candidate appearing inside, before, or after Si in the same window, it is predicted as a TR-opinion and its TA-target is the nearest tourist attraction.

***Table 4. Definitions of Base Rules***

| Rule | TR-Opinion Condition | TA-Target |
|---|---|---|
| *R*nt1 | $\exists ax$, $a \in \boldsymbol{TA_c}$ and $x \in ref_c(a)$ and $lst(x, 1, i) \geq 1$ | $\arg\max_{a \in TA_c, x \in ref_c(a)} lst(x,1,i)$ |
| *R*nt2 | $\exists ax$, $a \in \boldsymbol{TA_c}$ and $x \in ref_c(a)$ and $lst(x, i-b, i) \geq 1$ | $\arg\max_{a \in TA_c, x \in ref_c(a)} lst(x,i-b,i)$ |
| | $\exists ax$, $a \in \boldsymbol{TA_c}$ and $x \in ref_c(a)$ and $fst(x, i, i+b) \leq n$ | $\arg\min_{a \in TA_c, x \in ref_c(a)} fst(x,i,i+b)$ |

The choice of $\boldsymbol{S_c}$, $\boldsymbol{TA_c}$, and $ref_c(a)$ in *R*nt1 and *R*nt2 defines different rules to detect TR-opinions and TA-targets. These settings are quickly demonstrated in Table 4 and described more clearly in the following paragraphs.

**Baselines**

The baseline systems use the simplest way to make the first three decisions: (1) $\boldsymbol{S_c} = \boldsymbol{S_{op}}$, *i.e.* all of the opinionated sentences are TR-opinion candidates; (2) $\boldsymbol{TA_c} = \boldsymbol{TA}$, *i.e.* all of the tourist attractions in $D$ are TA-target candidates; and (3) $ref_c(a) = \{a\}$, *i.e.* only the full name of a tourist attraction is considered as a coreferential expression.

***Table 5. Rule Settings***

| Rule | Setting |
|---|---|
| Baselines | $\boldsymbol{S_c} = \boldsymbol{S_{op}}$, $\boldsymbol{TA_c} = \boldsymbol{TA}$, $ref_c(a) = \{a\}$ |
| *R*ow | $\boldsymbol{S_c} = \{S_i \mid S_i \in \boldsymbol{S_{op}}$ and $\exists x$, $x \in \boldsymbol{OW}$ and $in(x, i, i)=1\}$ |
| *R*mf | $\boldsymbol{TA_c}=\boldsymbol{A_{maxf}}$ |
| *R*cf | $ref_c(a) = ref_{all}(a)$ |

**TR-Opword Rule (*R*ow):**

In order to filter non-tourism-related sentences, such as bloggers' sentiments, an opinionated sentence is considered as a TR-opinion candidate only if it contains a TR-opword. The selection of $\boldsymbol{S_c}$ is given in the second row of Table 5.

**Most Frequent Tourist Attraction Rule (*R*mf)**

The most frequent tourist attraction appearing in a document *D* may be the focus of *D*. Many TR-opinions will target this tourist attraction. So, we only choose the most frequent tourist attractions in an article as the TA-target candidates, *i.e*. $\boldsymbol{TA_c = A_{maxf}}$.

**Coreferential Expression Rule (*R*cr)**

All kinds of coreferential expressions, as stated in Section 2.4, are considered when determining the occurrences of a tourist attraction *a*, *i.e*. $ref_c(a) = ref_{all}(a)$.

## 4. Machine Learning Approach

Approaches to build a TR-opinion analysis system by machine learning are described in this section. Such a system takes a whole article (including opinions and non-opinions) as its input and returns a set of TR-opinions together with their TA-targets. Features can be divided into two sets, which are defined in Section 4.1 and Section 4.2. The options of the system's architecture and training techniques are discussed in Section 4.3 and Section 4.4.

### 4.1 Features for TR-Opinion Detection

The first set of features is used to detect TR-opinions, *i.e*. to determine whether an opinionated sentence $S_i$ is tourism-related. Therefore, these features are designed for an opinionated sentence $S_i$. These features are quickly demonstrated in Table 6 and described more clearly in the following paragraphs.

**First Sentence Feature (*f*fs)**

The first sentence in an article often states the overall opinion of the author. It is interesting to see if the first sentence is tourism-related. The feature ***f*fs** finds the first sentence.

**TR-Opword Features (*f*ow$_{all}$ and *f*ow$_k$)**

If $S_i$ contains a TR-opword, it is likely to be a TR-opinion. Based on this idea, two kinds of features are defined: ***f*ow$_{all}$** checks if $S_i$ contains a TR-opword and ***f*ow$_k$** checks if $S_i$ contains a specific TR-opword $ow_k$.

***Table 6. Definition of TR-Opinion Detection Features***

| Feature | Definition of *feature*($S_i$) |
|---|---|
| *f*fs | 1 for $S_1$; 0 for other sentences in *D* |
| *f*ow$_{all}$ | 1 if $\exists x$, $x \in \boldsymbol{OW}$ and $in(x, i, i) = 1$; 0 otherwise |
| *f*ow$_k$ | 1 if $in(ow_k, i, i) = 1$; 0 otherwise |
| *f*ta$_{-1}$ / *f*tac$_{-1}$ | 1 if $\exists ax$, $[a \in \boldsymbol{TA}$ and $x \in ref_c(a)$ and $in(x, i-1, i-1) = 1]$; 0 otherwise |
| *f*ta$_0$ / *f*tac$_0$ | 1 if $\exists ax$, $[a \in \boldsymbol{TA}$ and $x \in ref_c(a)$ and $in(x, i, i) = 1]$; 0 otherwise |
| *f*ta$_{+1}$ / *f*tac$_{+1}$ | 1 if $\exists ax$, $[a \in \boldsymbol{TA}$ and $x \in ref_c(a)$ and $in(x, i+1, i+1) = 1]$; 0 otherwise |
| *f*ta$_{d-}$ / *f*tac$_{d-}$ | $1 - (i - Nid_{-}(S_i))/n$ |
| *f*ta$_{d+}$ / *f*tac$_{d+}$ | $1 - (Nid_{+}(S_i) - i)/n$ |
| *f*op$_{-1}$ | 1 if $Nop_{-}(S_i) = i-1$; 0 otherwise |
| *f*op$_{+1}$ | 1 if $Nop_{+}(S_i) = i+1$; 0 otherwise |
| *f*op$_{d-}$ | $1 - (i - Nop_{-}(S_i))/n$ |
| *f*op$_{d+}$ | $1 - (Nop_{+}(S_i) - i)/n$ |
| *f*to$_{-1}$ | 1 if the sentence preceding $S_i$ is a TR-opinion; 0 otherwise |
| *f*to$_{d-}$ | the distance score of the nearest TR-opinion preceding $S_i$ |
| *f*to$^{\#}$ | the 2 *f*to features whose values are assigned correctly |
| *f*to$^2$ | the 2 *f*to features whose values are predicted by a retrained classifier |

**Tourist Attraction Distance Feature (*f*ta and *f*tac)**

If an opinionated sentence is close to a tourist attraction, it is likely to be a TR-opinion and target that tourist attraction. Based on this idea, ten features are developed. The first five ***f*ta** features only consider full-name coreference, *i.e*. $ref_c(a) = \{a\}$:

***f*ta$_{-1}$**: check if the sentence preceding $S_i$ contains a tourist attraction

***f*ta$_0$**: check if $S_i$ contains a tourist attraction

***f*ta$_{+1}$**: check if the sentence following $S_i$ contains a tourist attraction

***f*ta$_{d-}$**: the distance score of the nearest tourist attraction preceding $S_i$

***f*ta$_{d+}$**: the distance score of the nearest tourist attraction following $S_i$

The next five features, ***f*tac$_{-1}$**, ***f*tac$_0$**, ***f*tac$_{+1}$**, ***f*tac$_{d-}$**, ***f*tac$_{d+}$**, are defined as the same as the five ***f*ta** features, except the choice of coreference can use all kinds coreferential expressions, *i.e*. $ref_c(a) = ref_{all}(a)$.

**Opinion Context Feature (*f*op)**

Four features come from the surrounding opinionated sentences.

***f*op$_{-1}$**: check if the sentence preceding $S_i$ is an opinion

***f*op$_{+1}$**: check if the sentence following $S_i$ is an opinion

***f*op$_{d-}$**: the distance score of the nearest opinion preceding $S_i$

***f*op$_{d+}$**: the distance score of the nearest opinion following $S_i$

**TR-Opinion Context Feature (*f*to)**

If an opinionated sentence is close to a TR-opinion, it is likely to be tourist-related, as well. Two features are introduced here:

***f*to$_{-1}$**: the sentence preceding $S_i$ is a TR-opinion

***f*to$_{d-}$**: the distance score of the nearest TR-opinion preceding $S_i$

Note that we do not know the values of these two features for a new article (nor should we when testing on the test set). In such a case, both feature values of the first sentence are set to be 0 because there is no preceding sentence. The predicted result of a sentence will be used to determine the two feature values of its following sentence. More ideas about these features are discussed in Section 4.4.

## 4.2 Features for TR-Target Identification

The second set of features is used to identify TA-targets, *i.e*. to determine whether a tourist attraction $A_j$ is the TA-target of an opinionated sentence $S_i$. Therefore, these features are designed for a pair of $<S_i, A_j>$ given an opinionated sentence $S_i$ and a tourist attraction $A_j$. These features are quickly demonstrated in Table 7 and described more clearly in the following paragraphs. The candidates of TA-targets are the set of tourist attractions appearing in the article.

**Frequency Feature (*f*fq)**

Similar to the idea of the Most-Frequent-Tourist-Attraction Rule, the occurrence of a tourist attraction is taken into account.

***Table 7. Definition of TR-Opinion Detection Features***

| **Feature** | **Definition of *feature*($S_i$, $A_j$)** |
|---|---|
| *f*fq | $freq(A_j)$ |
| *f*na$_{n-}$ / *f*nac$_{n-}$ | 1 if $Nta_{-}(S_i) = A_j$; 0 otherwise |
| *f*na$_{n+}$ / *f*nac$_{n+}$ | 1 if $Nta_{+}(S_i) = A_j$; 0 otherwise |
| *f*na$_{d-}$ / *f*nac$_{d-}$ | $1 - (i - Sid_{-}(A_j, S_i))/n$ |
| *f*na$_{d+}$ / *f*nac$_{d+}$ | $1 - (Sid_{+}(A_j, S_i) - i)/n$ |

**Distance Feature (*f*na and *f*nac)**

It is intuitive that a TR-opinion is often close to its targeting tourist attraction. Eight features are derived from the distance of an opinionated sentence $S_i$ and a tourist attraction $A_j$. The first four ***f*na** features only consider full-name coreference, *i.e.* $ref_c(a) = \{a\}$:

$\boldsymbol{fna_{n-}}$: check if $A_j$ is the nearest tourist attraction preceding $S_i$

$\boldsymbol{fna_{n+}}$: check if $A_j$ is the nearest tourist attraction following $S_i$

$\boldsymbol{fna_{d-}}$: the distance score of $A_j$ and $S_i$ when $A_j$ precedes $S_i$

$\boldsymbol{fna_{d+}}$: the distance score of $A_j$ and $S_i$ when $A_j$ follows $S_i$

The next four features, $\boldsymbol{fnac_{n-}}$, $\boldsymbol{fnac_{n+}}$, $\boldsymbol{fnac_{d-}}$, $\boldsymbol{fnac_{d+}}$, are defined as the same as the four ***f*na** features, except the choice of coreference can use all kinds coreferential expressions, *i.e.* $ref_c(a) = ref_{all}(a)$.

## 4.3 Retraining by Prediction

The TR-Opinion Context Feature (***f*to**) is very useful but also dangerous. We conducted an oracle model where the values of the TR-Opinion Context Feature of the test data were set correctly (denoted as $\boldsymbol{fto^{\#}}$), and found that the performance was the best (as depicted later in Table 10). Nevertheless, if the feature values came from the predictions of the classifier, the errors would propagate and harm the performance greatly (also depicted in Table 10).

We propose a retraining method to use the TR-Opinion Context Feature. Training is performed in three steps. First, set the values of the TR-Opinion Context Feature of the training data correctly to train a preliminary classifier. Use this preliminary classifier to predict the TR-opinions in the training set. Then, use the predictions to assign the values of the TR-Opinion Context Feature of the training data to train a classifier. The second classifier is used to construct the real TA-target identification system. The values of the TR-Opinion Context Feature predicted by the second classifier are denoted as $\boldsymbol{fto^{2}}$.

## 4.4 Single-Layer and Dual-Layer Models

Our TA-target identification system is constructed as follows: each sentence in an article is paired with each of the tourist attractions appearing in the article and labeled by a classifier. If none of the pairs is classified as positive, this sentence is not a TR-opinion. Otherwise, the sentence is predicted as a TR-opinion and all the tourist attractions in the pairs receiving positive predictions are its TR-targets.

The process of TA-target identification can be divided into two steps: detecting TR-opinions and assigning TR-targets to them. Hence, we can train two classifiers for the two steps separately, or train a single classifier to identify the TA-targets directly. Two different models are designed, given that the input is a pair of an opinionated sentence $S_i$ and a tourist attraction $A_j$.

**Single-Layer Model**

The classifier directly determines whether the tourist attraction $A_j$ is the TR-target of the sentence $S_i$. All of the features introduced in Section 4.1 and 4.2 are used for training even if a feature only relates to the sentence $S_i$ only.

**Dual-Layer Model**

The classification module consists of two classifiers. The first-layer classifier determines whether $S_i$ is a TR-opinion. Only features introduced in Section 4.1 are used to train the first-layer classifier. If $S_i$ is classified as a TR-opinion, the pair $<S_i, A_j>$ is passed to the second-layer classifier. The second-layer classifier determines whether $A_j$ is the TR-target of $S_i$. Only features introduced in Section 4.2 are used to train the second-layer classifier.

## 5. Experiments

The experiments shown in this section were all conducted in a leave-one-out cross-validation fashion where each of the 156 articles in the experimental data set was kept out as the test data and the others as the training data in turn.

The number of the positive examples is relatively small compared to the negative examples. We did not evaluate the system by accuracy because the majority prefers guessing all sentences as “not TR-opinion”. Additionally, in order to create a balanced training set, we randomly selected negative examples in the same amount of the positive examples in each training set.

Both TR-opinion detection and TA-target identification are evaluated by the micro-average precision (P), recall (R), and F-measure (F), where $F = \frac{2 \times P \times R}{P + R}$.

For TR-opinion detection,

$$P = \frac{\#(\text{correctly guessed TR - opinions})}{\#(\text{TR - opinions guessed by system})} \quad (2)$$

$$R = \frac{\#(\text{correctly guessed TR - opinions})}{\#(\text{real TR - opinions})} \quad (3)$$

For TA-target identification,

$$P = \frac{\#(\text{correctly guessed TA - targets})}{\#(\text{TA - targets guessed by system})} \quad (4)$$

$$R = \frac{\#(\text{correctly guessed TA - targets})}{\#(\text{real TA - targets})} \quad (5)$$

## 5.1 Tourism-Related Opinion Word Selection

As introduced in Section 2.3, we want to find opinion words highly related to tourism. A preliminary experiment was conducted to determine the threshold of TR-scores to select TR-opwords. The candidates of TR-opwords were the opinion words collected in NTUSD, the National Taiwan University Sentiment Dictionary (Ku & Chen, 2007).

The threshold of the TR-scores was determined by the baseline experiment of TR-opinion detection. Set the threshold values varying from 0 to 1 with a step of 0.01 and selected those opinion words whose TR-scores were higher than the threshold to predict TR-opinions by the TR-Opword Rule only.

***Table 8. Performance of TR-Opinion Detection under Different Thresholds***

| **Threshold** | **#TR-ow** | **P** | **R** | **F** |
|---|---|---|---|---|
| 0 | 482.1 | 37.71 | 46.46 | 41.63 |
| 0.1 | 475.2 | 38.71 | 46.04 | 42.06 |
| 0.2 | 443.5 | 41.42 | 43.29 | 42.33 |
| **0.25** | **418.6** | **43.17** | **41.62** | **42.38** |
| **0.26** | **418.6** | **43.17** | **41.62** | **42.38** |
| 0.3 | 408.8 | 42.82 | 39.78 | 41.25 |
| 0.4 | 359.7 | 46.58 | 31.78 | 37.78 |
| 0.5 | 266.2 | 49.28 | 22.77 | 31.15 |
| 0.6 | 251.3 | 50.23 | 18.18 | 26.70 |
| 0.7 | 218.4 | 49.06 | 10.93 | 17.87 |
| 0.8 | 202.5 | 50.50 | 8.42 | 14.44 |

Table 8 shows the results of TR-opinion detection under different threshold settings. The threshold value achieving the best performance was 0.25 and 0.26, but not significantly the best if compared to a nearby setting. We chose 0.25 as the threshold in the following experiments. Note that the sets of TR-opwords were not the same in different iterations of cross-validation because the training sets were different. The second column of Table 8 depicts the average number of TR-opwords selected in each iteration.

## 5.2 Experiments of Rule-Based Approaches

Table 9 presents the results of the rule-based TA-target identification systems under different rule combinations. The Nearest-TA-in-Window Rule (*R*nt2) slightly outperformed the Nearest- Preceding-TA Rule (*R*nt1) in any combination. The rule combination achieving the best performance was the Nearest-TA-in-Window Rule (*R*nt2) combined with the Coreferential Expression Rule (*R*cr), which was significantly different from all the others.

***Table 9. Performance of the Rule-Based TA-Target Identification Systems***

| **Rule Combination** | **P** | **R** | **F** |
|---|---|---|---|
| *R*nt1 | 25.74 | 70.73 | 37.74 |
| *R*nt1+*R*ow | 32.21 | 29.44 | 30.76 |
| *R*nt1+*R*mf | 18.84 | 46.96 | 26.89 |
| *R*nt1+*R*cr | 27.01 | 74.65 | 39.67 |
| *R*nt1+*R*ow+*R*cr | 19.16 | 47.79 | 27.35 |
| *R*nt1+*R*mf+*R*cr | 34.18 | 31.28 | 32.67 |
| *R*nt1+*R*ow+*R*mf+*R*cr | 23.16 | 19.43 | 21.13 |
| *R*nt2 (*b*=5) | 29.93 | 52.54 | 38.14 |
| *R*nt2+*R*ow | 35.21 | 21.93 | 27.03 |
| *R*nt2+*R*mf | 22.90 | 26.61 | 24.61 |
| ***R*nt2+*R*cr** | **32.10** | **60.88** | **42.04** |
| *R*nt2+*R*ow+*R*cr | 25.34 | 31.53 | 28.09 |
| *R*nt2+*R*mf+*R*cr | 37.47 | 25.19 | 30.12 |
| *R*nt2+*R*ow+*R*mf+*R*cr | 28.46 | 12.68 | 17.54 |

## 5.3 Experiments of Machine Learning Approaches

We used the LIBSVM tool (Fan *et al*., 2005) to train the classifiers. We chose SVM because some features' domains were sets of real numbers, not strings.

The dual-layer model first detects the TR-opinions then identifies the TA-targets. We evaluated the first-layer (for TR-opinion detection) and second-layer (for TA-target identification) classifiers separately.

### 5.3.1 TR-Opinion Detection Experiments

Table 10 presents the selected results of TR-opinion detection by different combinations of features where $f\mathrm{xx}_{-}$ denotes all $f$xx features regarding objects preceding the sentence (*i.e*. $f\mathrm{xx}_{-1}$ and $f\mathrm{xx}_{d-}$), and $f\mathrm{xx}_{0-}$ denotes the feature combination of $f\mathrm{xx}_{-}$ and $f\mathrm{xx}_{0}$.

The results in Table 10 are represented in groups. The experiments in the first group only used the Tourist Attraction Distance Features ($f$ta). The feature combinations in the second group were suggested by a feature selection method, WLLR, which will be introduced later.

***Table 10. Results of the TR-Opinion Detection by Machine Learning, Rules, and Annotators***

| Feature Combination | P | R | F |
|---|---|---|---|
| $f\text{ta}$ | 42.15 | 60.88 | 49.81 |
| $f\text{ta}_{-}$ | 40.92 | 80.23 | 54.20 |
| $f\text{ta}_{0-}$ | 61.18 | 36.28 | 45.55 |
| $f\text{tac}$ | 56.90 | 47.79 | 51.95 |
| $f\text{tac}_{-}$ | 41.95 | 84.07 | 55.97 |
| $f\text{tac}_{0-}$ | 62.28 | 44.20 | 51.71 |
| $f\text{ow}_{all}+f\text{tac}+f\text{to}^2$ | 55.67 | 58.97 | 57.27 |
| $f\text{ow}_{all}+f\text{tac}_{0-}+f\text{to}^2$ | 54.91 | 60.13 | 57.40 |
| $f\text{ow}_{all}+f\text{fs}+f\text{op}_{-}+f\text{tac}_{-}+f\text{to}^2$ | 48.48 | 61.38 | 54.18 |
| $f\text{ow}_{all}+f\text{fs}+f\text{op}_{-}+f\text{tac}_{0-}+f\text{to}^2$ | 54.34 | 58.97 | 56.56 |
| $\boldsymbol{f\text{ow}_{all}+f\text{fs}+f\text{op}+f\text{tac}+f\text{to}^2}$ | **55.98** | **59.30** | **57.59** |
| $f\text{ow}_{all}+f\text{fs}+f\text{op}+f\text{ta}+f\text{to}^2$ | 50.68 | 53.13 | 51.87 |
| $f\text{ow}_{all}+f\text{fs}+f\text{op}_{-}+f\text{to}^{\#}$ | 58.77 | 79.40 | 67.54 |
| $f\text{ow}_{all}+f\text{fs}+f\text{op}_{-}+f\text{tac}_{-}+f\text{to}^{\#}$ | 65.37 | 64.22 | 64.79 |
| $f\text{ow}_{all}+f\text{fs}+f\text{op}_{-}+f\text{tac}_{-}+f\text{to}$ | 57.60 | 40.12 | 47.30 |
| $R\text{nt2}+R\text{cr}$ | 43.14 | 81.82 | 56.49 |
| Annotator 1 | 85.62 | 88.91 | 87.23 |
| Annotator 2 | 89.17 | 82.40 | 85.65 |
| Annotator 3 | 96.52 | 57.80 | 72.30 |

The experiments in the third and the fourth groups tried more feature combinations but used the TR-opinion Context Features in different ways. The fourth group used the TR-opinion Context Feature after Retraining ($f\text{to}^2$). The fourth group used correct values for the TR-opinion Context Features ($f\text{to}^{\#}$, as oracle model) and prediction by the previously trained model without retraining ($f\text{to}$).

The fifth one has the best performance achieved by the rule-based model and the final group lists the performances of human annotators which can be regarded as upper bounds.

The second and the third groups of results show that the TR-opinion Context Feature after Retraining ($f\text{to}^2$) is useful, for the best performances were achieved by those feature combinations containing $f\text{to}^2$. Compared with the fourth group, the oracle model (containing

$f$to$^{\#}$) outperforms other combinations, which concludes that $f$to$^{\#}$ is a great feature but, unfortunately, is unattainable. On the other hand, using the prediction by the classifier without retraining ($f$to) harmed the performance. We can say that the retraining process did improve the performance.

The first group also suggests that the Preceding Tourist Attraction Distance Features with or without Coreferential Expressions ($f$ta$_{-}$ and $f$tac$_{-}$) are useful.

To see the usefulness of features, we used an adapted version of WLLR (Weighted Log Likelihood Ratio) (Nigam *et al*., 2000) to measure the usefulness of the features. The adapted equation of WLLR in our work is:

$$WLLR(f) = \underset{x \in P}{avg}(f(x)) \log \frac{\underset{x \in P}{avg}(f(x))}{\underset{x \in N}{avg}(f(x))} \tag{6}$$

***Table 11. WLLR of Features***

| **Feature** | $avg_P(f)$ | $avg_P(f) / avg_N(f)$ | **WLLR** |
|---|---|---|---|
| $f$to$^{\#}_{-1}$ | 0.371 | 8.204 | 0.781 |
| $f$tac$_0$ | 0.272 | 5.588 | 0.468 |
| $f$to$^{\#}_{d-}$ | 0.853 | 1.599 | 0.401 |
| $f$ta$_0$ | 0.220 | 5.930 | 0.392 |
| $f$tac$_{-1}$ | 0.258 | 2.614 | 0.248 |
| $f$tac$_{d+}$ | 0.832 | 1.280 | 0.205 |
| $f$ta$_{-1}$ | 0.210 | 2.438 | 0.187 |
| $f$ta$_{d+}$ | 0.788 | 1.259 | 0.181 |
| $f$ow$_{all}$ | 0.416 | 1.484 | 0.164 |
| $f$tac$_{d-}$ | 0.903 | 1.198 | 0.163 |
| $f$ta$_{d-}$ | 0.875 | 1.185 | 0.148 |
| $f$tac$_{+1}$ | 0.192 | 1.677 | 0.099 |
| $f$ta$_{+1}$ | 0.160 | 1.638 | 0.079 |
| $f$op$_{d+}$ | 0.938 | 1.028 | 0.026 |
| $f$op$_{d-}$ | 0.931 | 1.017 | 0.015 |
| $f$op$_{-1}$ | 0.463 | 1.033 | 0.015 |
| $f$op$_{+1}$ | 0.460 | 1.022 | 0.010 |
| $f$fs | 0.038 | 0.817 | -0.008 |

where $f(x)$ is a feature function which defines a numerical feature value for a given example $x$, $avg(\boldsymbol{v})$ means the average over a numerical set $\boldsymbol{v}$, $P$ and $N$ are the sets of positive examples and negative examples in the training set, respectively. The adaptation is made to make it applicable for both Boolean features (treated as 0 and 1) and numerical features.

Table 11 lists the WLLR and averages (over positive and negative examples) of the features. As we can see, the best features according to WLLR are the TR-Opinion Context Features (***f*to**), the Tourist Attraction Distance Features (***f*ta** and ***f*tac**, with or without coreferential expressions), and the All-TR-Opword Feature ($\boldsymbol{f}\mathbf{ow_{all}}$). The experiments inspired by feature selection are listed in the second group. The results in Table 10 support the predictions by WLLR as the feature combination $f\text{ow}_{all}+f\text{tac}_0+f\text{to}^2$ performs very well.

The best performance, however, where an F-measure score of 57.59% is achieved, is by the feature combination using all kinds of features. It outperforms the combination by feature selection significantly ($p<0.001$).

### 5.3.2 TA-Target Identification Experiments

Table 12 lists the experimental results of TA-target identification by different approaches. The second row gives the performance of the second-layer classifier where the first-layer was replaced by a perfect model, *i.e.* only known TR-opinions were assigned TA-targets. The precision and recall scores were 90.06% and 89.91%, respectively, and the F-measure score was around 90%. This means that the bottleneck of this work is TR-opinion detection. The third row shows the performance of the overall dual-layer system consisting of the best models of the two layers, which F-measure is 52.72% and is the best among all TA-target identification models.

The models of the fourth and the fifth rows are single-layer classifiers. Even when the correct values of TR-Opinion Context Features ($f\text{to}^{\#}$) are used, they still cannot compete with the dual-layer model. This shows that dual-layer classification is a better approach.

The sixth row of Table 12 gives the performance of TA-target identification by rules. Although the best rule-based approach performs well in TR-opinion detection, its ability to identify TA-targets is weaker.

The last three rows present the performance of the results of the three annotators. We can see that the best F-measure of a ML-based system is about 60% to 75% of human ability. So, there is still room to improve.

***Table 12. Results of TA-Target Identification by Different Approaches***

| **Feature Combination** | **P** | **R** | **F** |
|---|---|---|---|
| **The second layer only (TA-Target Identification)** | | | |
| *f*fq+*f*nac | 90.06 | 89.91 | 89.98 |
| **Dual-Layer Model** | | | |
| 1st layer: *f*ow$_{all}$+*f*fs+*f*op+*f*tac+*f*to$^2$<br>2nd layer: *f*fq+*f*nac | **51.30** | **54.21** | **52.72** |
| **Single-Layer Model** | | | |
| *f*ow$_{all}$+*f*fs+*f*op$_-$+*f*to$^\#$+*f*fq+*f*nac | 32.83 | 88.91 | 47.95 |
| *f*ow$_{all}$+*f*fs+*f*op$_-$+*f*tac+*f*to$^\#$+*f*fq+*f*nac | 32.75 | 88.74 | 47.84 |
| *R*nt2+*R*cr | 32.10 | 60.88 | 42.04 |
| Annotator 1 | 84.10 | 87.32 | 85.68 |
| Annotator 2 | 87.27 | 80.65 | 83.83 |
| Annotator 3 | 94.71 | 56.71 | 70.94 |

## 6. Conclusions and Future Work

This paper aims at detecting tourism-related opinionated sentences and identifying their tourist attraction targets. Several rules and features were invented and tested in different combinations. The performance is improved by building a dual-layer classification system where the classifiers of TR-opinion detection and TA-target identification are trained separately. Retraining by the prediction method is introduced to decide the values of the TR-Opinion Context Features. This feature, together with the tourism-related opinion words and distances to the tourist attractions were verified to be useful. The best overall performance of TA-target identification is 52.72%, which is about 60% to 75% of human ability.

In the future, we would like to implement known methods to do opinion detection and tourist attraction recognition so we can build a real system and evaluate its performance. More features should be studied for TR-opinion detection.

By the location information of the tourist attractions, it is also interesting to make a summary for a city or a country by the opinions about the tourist attractions located in that area. This will be our future work.

## Reference


Alm, C. O., Roth, D., & Sproat, R. (2005). Emotions from text: machine learning for text-based emotion prediction. In *Proceeding of HLT/EMNLP 2005*, 579-586.

Breck, E., Choi, Y., & Cardie, C. (2007). Identifying Expressions of Opinion in Context. In *Proceeding of IJCAI 2007*, 2683-2688.

Choi, Y., Cardie, C., Riloff, E., & Patwardhan, S. (2005). Identifying Sources of Opinions with Conditional Random Fields and Extraction Patterns. In *Proceeding of HLT/EMNLP 2005*, 355-362.

Fan, R.E., Chen, P.H., Lin, C.J., & Joachims, T. (2005). Working Set Selection Using the Second Order Information for Training SVM. In *Journal of Machine Learning Research*, 6, 1889-1918.

Ghani, R., Probst, K., Liu, Y., Krema, M., & Fano, A. (2006). Text Mining for Product Attribute Extraction. In *SIGKDD Explorations*, 1(8), 41-48.

Ghose, A., Ipeirotis, P., & Sundararajan, A. (2007). Opinion Mining using Econometrics: A Case Study on Reputation. In *Proceeding of ACL 2007*, 416-423.

Godbole, N., Srinivasaiah, M., & Skiena, S. (2007). LargeScale Sentiment Analysis for News and Blogs. In *Proceedings of ICWSM 2007*, 219-222.

Hu M. & Liu, B. (2004). Mining Opinion Features in Customer Reviews. In *Proceeding of AAAI 2004*, 755-760.

Kim, S.M., & Hovy, E. (2005). Identifying Opinion Holders for Question Answering in Opinion Texts. In *Proceedings of AAAI-05 Workshop on Question Answering in Restricted Domains*, 1-8.

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

Ku, L.W., Lee, L.Y., Wu, T.H., & Chen, H.H. (2005). Major Topic Detection and Its Application to Opinion Summarization. In *Proceedings of SIGIR 2005*, 627-628.

Nigam, K., McCallum, A. K., Thrun, S., & Mitchell, T. (2000). Text Classification from Labeled and Unlabeled Documents Using EM. In *Machine Learning*, 39(2-3), 103-134.

Okamoto, T., Honda, T., & Eguchi, K. (2009). Locally Contextualized Smoothing of Language Models for Sentiment Sentence Retrieval. In *Proceeding of the 1st international CIKM 2009 Workshop on Topic-Sentiment Analysis for Mass Opinion*, 73-80.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification Using Machine Learning Techniques. In *Proceeding of EMNLP 2002*, 79-86.

Seki, Y., Ku, L.W., Sun, L., Chen, H.H., & Kando, N. (2010). Overview of Multilingual Opinion Analysis Task at NTCIR-8: A Step Toward Cross Lingual Opinion Analysis. In *Proceedings of NTCIR-8*, 209-220.

Wiebe, J., Wilson, T., & Bell, M. (2001). Identifying Collocations for Recognizing Opinions. In *Proceeding of ACL 2001 Workshop on Collocation*, 24-31.

Xia, Y., Hao, B., & Wong, K.F. (2009). Opinion Target Network and Bootstrapping Method for Chinese Opinion Target Extraction. In *Lecture Notes in Computer Science*, 5839, 339-350.

Zhang, W., Jia, L., Yu, C., & Meng, W. (2008). Improve the Effectiveness of the Opinion Retrieval and Opinion Polarity Classification. In *Proceedings of the CIKM 2008*, 1415-1416.