**City University of Hong Kong**

# Aspect-oriented Sentiment Analysis on Video Game Reviews in Chinese

By

LEUNG Hoi Tik

May 2010

Submitted to the Department of Chinese, Translation and Linguistics

in partial fulfillment of the requirements for the degree of

Bachelor of Arts (Honours) in Linguistics and Language Technology

## Abstract

This project develops a system for aspect-oriented sentiment analysis on video game reviews in Chinese. The system focuses on the major tasks of sentiment analysis: identification of subjective linguistic units and targets of sentiments, evaluation of the polarity of sentiments and summarization of sentiments. The proximity between intensifiers and opinion words and that between opinion words and keywords of aspects of video games are major clues to handle the tasks. The output of the system is a summary of a reviewer's opinions on the relative quality of various aspects of a reviewed game. Despite the fair performance of the system, future directions for its improvement are identified and discussed.



**Keywords:** Chinese video game reviews, opinion, polarity, sentiment analysis

## Table of Contents

**Abstract** *i*

**Table of Contents** *ii*

**1 Introduction** *1*

**1.1 Project Theme and Motivations** *1*

**1.2 Objectives** *1*

**2 Related Work** *2*

**2.1 Identification of Subjective Linguistic Units** *2*

**2.2 Evaluation of the Polarity of Sentiments** *4*

**2.3 Identification of Holders and Targets of Sentiments** *5*

**2.4 Summarization of Sentiments** *6*

**2.5 Other Relevant Issues** *7*

**2.6 Approaches of the Current Project** *8*

**2.6.1 Identification of Subjective Linguistic Units and Targets of Sentiments** *8*

**2.6.2 Evaluation of the Polarity of Sentiments** *8*

**2.6.3 Summarization of Sentiments** *8*

**3 System Description** *8*

**3.1 Assumptions on Input Reviews** *9*

**3.2 Identification of Subjective Linguistic Units and Targets of Sentiments** *9*

**3.3 Evaluation of the Polarity of Sentiments** *10*

**3.4 Summarization of Sentiments** *10*

**4 Data and Resources** *11*

**4.1 Collection of Sample Reviews** *11*

**4.2 Collection of Words for Compiling Dictionaries** *12*

**4.2.1 Format for Storing Intensifiers** *12*

**4.2.2 Format for Storing Opinion Words** *12*

**4.2.3 Format for Storing Keywords of Aspects** *12*

**4.2.4 Extraction of Target Words from Reviews** *12*
**4.2.5 Judgement on Attributes of Identified Words** *13*

**5 Program Implementation** *14*
**5.1 Input Handling** *14*
**5.2 Identification of Subjective Linguistic Units and Targets of Sentiments** *15*
**5.3 Evaluation of the Polarity of Sentiments** *15*
**5.3.1 Calculation of Intensity Values with Polarity of Opinion Words** *15*
**5.3.2 Calculation of Overall Sentiment Scores of Aspects** *16*
**5.4 Summarization of Sentiments** *17*

**6 System Results and Discussion** *18*
**6.1 Preparation of Gold-standard Outputs** *18*
**6.2 Test Result** *19*
**6.3 Discussion** *19*
**6.3.1 Failure in Identification of Implicit Opinions** *19*
**6.3.2 Insufficient Information in Dictionaries** *19*
**6.3.3 Discrepancy on Intensity Values** *20*
**6.3.4 Algorithm-driven Problems** *20*

**7 Future Work** *21*

**8 Conclusion** *22*

**References** *23*

**Appendix: Sample Comment Sheet** *25*

# 1 Introduction

In recent years, computational subjectivity and sentiment analyses have drawn much attention in various domains, such as Web documents, product reviews, news articles, etc. (Kim & Hovy, 2006; Kobayashi, Inui, & Matsumoto, 2007; Liu, Cao, Lin, Huang, & Zhou, 2007; Pang & Lee, 2004; Popescu & Etzioni, 2005). The major tasks in such analyses include identifying subjective linguistic units, such as words, phrases, etc., which express opinions or sentiments in a piece of raw text, evaluating the polarity of sentiments of those linguistic units, finding the holders and targets of the sentiments and summarizing the sentiments expressed in the text (Hatzivassiloglou & McKeown, 1997; Kim & Hovy, 2006; Meng & Wang, 2009; Popescu & Etzioni, 2005; Turney, 2002; Riloff, Wiebe, & Wilson, 2003). Many studies in this field have targeted on English text while the investigation on Chinese text has just sprouted (Tsou, Kwong, Wong, & Lai, 2005).

## 1.1 Project Theme and Motivations

This project focuses on developing a system for aspect-oriented sentiment analysis on video game reviews in Chinese. Traditionally, many video game players (gamers) made decisions to buy new games depending heavily on editors' reviews on newly released games in printed video game magazines until the generation of Web 2.0 emerged. Nowadays, video game reviews are shared on the Internet in different forms, such as videos, Web forum posts, weblogs, etc. The authors of these reviews are no longer restricted to a small group of professional game magazine editors but include also any parties who want to share their views on video games. This leads to the rapid expansion of video game reviews on the Internet. As a consequence, searching for desired information in excessively available game reviews becomes a time-consuming task. Thus, the development of a system for analyzing sentiments in video game reviews is motivated.

## 1.2 Objectives

This project aims to develop a system which analyzes the sentiments expressed by the author in an unannotated segmented video game review in Chinese. The output is expected to provide users a general summary of the quality of the video game reviewed. The summary shows the relative quality of four aspects of the game, namely System, Story, Image and Sound, expressed in the game review. Users can therefore know which aspects are relatively outstanding or less inferior among the various aspects. Take the following virtual input review as an example:

這遊戲的系統及畫面真的很好。音效及故事也不錯。
"The system of this game and images inside are really very good. Sound effects and the story are not bad."

The output of the system is expected to be a summary like the following one:

Four aspects of the game are commented by the author.
In descending order of quality, they are: System = Image > Story = Sound
N.B.:
The Equality sign (=) indicates that the preceding aspect is of the same quality as the succeeding aspect.
The Greater-than sign (>) indicates that the preceding aspect is of higher quality than the succeeding aspect.

The focus is placed on extracting sentiment-bearing words and computing the polarity of sentiments which belong to different aspects. Other necessary procedures, such as word segmentation of the input text, correction of typing mistakes in the input text, etc., are either done manually or assumed to have been done.

## 2 Related Work

Computational sentiment analysis involves a series of tasks as mentioned above. None of the tasks can be performed with no errors or mistakes. Therefore, to maximize the reliability of systems and the accuracy of results is a challenging mission for computational linguists. Many of them have addressed problems in one or more areas of sentiment analysis.

### 2.1 Identification of Subjective Linguistic Units

Wiebe, Bruce and O'Hara (1999), Turney (2002), Riloff et al. (2003), Pang and Lee (2004) and Popescu and Etzioni (2005) have addressed different issues related to the identification of subjective linguistic units.

Wiebe et al. (1999) studied the use of statistical method to improve manual subjectivity judgement. Their study focused on sentences in news reporting in which information might be subjective or objective. They addressed the problem of disagreement in subjectivity tagging caused by bias among different judges. Accordingly, bias-corrected tags were developed statistically and were used to revise the coding manual for judges and to develop an automatic classifier. The

revised coding manual improved the agreement among the judges. The good performance of the developed classifier proved that their study was valuable.

Turney (2002) employed a part-of-speech tagger to extract adjectives or adverbs as subjective words. He worked on reviews from different domains: automobiles, banks, movies and travel destinations. He noted that semantic orientation of a subjective word highly depended on context. Therefore, instead of extracting a single word, he proposed to extract a phrase of two consecutive words where one of them was the subjective word and the other provided context.

Riloff et al. (2003) adopted two bootstrapping algorithms to learn subjective nouns from unannotated text extracted from news documents. The learned subjective nouns were then used together with other features such as manually developed features and discourse features to train a subjectivity classifier which was used to identify subjective sentences. As a result, the performance of the classifier was improved by employing the learned subjective nouns as a feature. Unlike previous studies which had focused on exploring adjectival phrases (Turney, 2002), their study supplemented the area of subjectivity mining by investigating the influence of subjective nouns.

Pang and Lee (2004) presented a way to extract the subjective portions of the whole text for further analysis. They first employed graph-cut-based subjectivity detectors to distinguish and extract subjective sentences from objective sentences in the input movie reviews. Next, they evaluated the sentiments extracts to see whether they were positive or negative reviews. The subjectivity detectors were trained by a collection of sentences which were labelled as subjective or objective. Their work showed that the extracted portions still retained sufficient polarity information compared with the original full text. In addition, they suggested that the shorter extracts enhanced the effectiveness of polarity classification and further improvement was made by considering contextual information like sentence proximity. Their study also contributed to the efficiency of sentiment analysis.

Popescu and Etzioni (2005) introduced an unsupervised information-extraction system, OPINE, which was used to extract product features and associated opinions and to evaluate the polarity of the opinions in product reviews. For the extraction of opinions, OPINE relied on explicit clues to identify potential opinion phrases. First, extraction rules built on syntactic dependencies were applied to identify potential opinion phrases which consisted of head words and their modifiers. The relaxation labelling technique was employed to label the semantic orientation of the head words in the potential opinion phrases as positive, negative or neutral. Potential opinion phrases with

head words which were labelled as positive or negative were recognized as opinion phrases.

## 2.2 Evaluation of the Polarity of Sentiments

Hatzivassiloglou and McKeown (1997), Turney (2002), Popescu and Etzioni (2005), Tsou et al. (2005) and Min and Park (2009) have investigated the evaluation of the polarity of sentiments.

Hatzivassiloglou and McKeown (1997) adopted an automatic log-linear regression model to apply linguistic constraints from conjunctions to the classification of semantic orientation of conjoined adjectives. They suggested that most of the conjunctions implicitly provided information about the semantic orientation of the conjoined adjectives. On the one hand, most of them usually connected adjectives of the same orientation. On the other hand, conjunctions like "but" were likely to connect two adjectives of different orientations. In addition, they proposed that many morphologically related pairs such as "adequate-inadequate" had different orientations. Furthermore, they mentioned that the semantically unmarked member, which had a high frequency of occurrence, of a pair of gradable adjectives usually had positive orientation. After all, they used a clustering algorithm to separate conjoined adjectives into two groups of different orientations. The group had higher average frequency were then labelled as positive. Their study presented how surrounding contextual information could be exploited to evaluate the polarity of subjective units.

Turney (2002) employed a probabilistic algorithm to find out the strength of the association between the subjective phrase and reference words of semantic orientation, such as "excellent" and "poor". The phrase was identified as positive if it was associated more strongly with "excellent", otherwise, it was negative.

As mentioned in the above section, OPINE used a relaxation labelling technique to assign semantic orientation labels to words (Popescu & Etzioni, 2005). Relaxation labelling made use of probabilistic computation, word relationships and syntactic relationships to assign semantic orientation labels to words. In contrast to previous studies (Hatzivassiloglou & McKeown, 1997; Turney, 2002) which treated semantic orientation as a binary feature that was either positive or negative, Popescu and Etzioni (2005) improved the labelling of semantic orientation by adding neutral as one of the possible labels.

In addition to positive and negative polarity, Tsou et al. (2005) proposed three attributes of polar items. The first one was Spread, which was the ratio of the number of coherent textual segments (CTS) which carried polar elements to the total number of CTS in an article. The second one was Density, which meant

"how extensive polar elements are found in a polar CTS on average". The last one was Intensity, which indicated how strong or subjective an expression was. By proposing Spread, Density and Intensity, they provided different perspectives on the evaluation of polarity.

Min and Park (2009) proposed finer-grained categories of polarity to enhance the effectiveness of summarizing sentiments in product reviews. They pointed out the importance of the weight of polarity. If the weight was not taken into consideration and the overall sentiment was done by simple summation of the number of sentiments, some details would be missed. They further suggested that clues such as "relevance for satisfaction", "contrastive weight" and some adverbs played a crucial role in determining the polarity of sentiments

### 2.3 Identification of Holders and Targets of Sentiments

Kim and Hovy (2006) have contributed to the identification of holders and targets of sentiments while Popescu and Etzioni (2005), Tsou et al. (2005), Kobayashi et al. (2007) and Meng and Wang (2009) have studied the identification of targets of sentiments.

Kim and Hovy (2006) proposed the exploitation of the semantic structure of a sentence in identifying the holders and the topics or targets of sentiments. In their study, they made use of semantically annotated data retrieved from FrameNet[1] (Fillmore & Baker, n.d.). In FrameNet, different kinds of frames are assigned to words. According to Ruppenhofer, Ellsworth, Petruck, Johnson and Scheffczyk (2006), a frame is "a script-like conceptual structure that describes a particular type of situation, object, or event along with its participants and props". Kim and Hovy (2006) first selected suitable frames collected from FrameNet for identified opinion words. Next, they identified candidates of the frame elements in parsed sentences. After that, semantic roles were assigned to each frame element. In the end, holders and targets of the opinion words were determined among labelled semantic roles.

According to Popescu and Etzioni (2005), OPINE made use of parsed review data supplemented with statistical methods to extract product features, which were the targets of sentiments, from product reviews.

Tsou et al. (2005) addressed the problem that a coherent textual segment in an article might contain more than one focus. This posed difficulties for a system to identify the actual target of sentiments in such a segment and further work should be done to help identifying the targets.

Kobayashi et al. (2007) proposed a machine learning-based method for

[1] http://framenet.icsi.berkeley.edu/

extracting aspect-evaluation relation and aspect-of relation based on contextual and statistical clues. Contextual clues were syntactic patterns of sentences. Statistical clues were the statistics of co-occurrences of aspect-aspect and aspect-evaluation. An aspect was a part of an entity on which an evaluation was made. An evaluation was a phrase that expressed an opinion. Hence, for example, "curry with chicken" and "was good" in the passage "I went out for lunch at the Deli and ordered a curry with chicken. It was pretty good" formed an aspect-evaluation relation where "was good" was an evaluation made on "curry with chicken". Also, "the Deli" and "curry with the chicken" formed an aspect-of relation where "curry with the chicken" was an aspect (daughter) of the subject (parent) "the Deli" in the hierarchy of aspect-of relation. For extracting aspect-evaluation relation, Kobayashi et al. (2007) first used a dictionary of evaluation expressions to assist the identification of evaluations in weblog posts. Next, they identified the targets of those evaluations by exploiting contextual and statistical clues. After that, they proceeded to extract aspect-of relation. For extracting aspect-of relation, they also exploited contextual and statistical clues. If the identified aspects themselves were not the upmost subjects in the hierarchy of aspect-of relation, they recursively searched for the parent aspects until the upmost aspects were found. In contrast to the approach of Kim and Hovy (2006), by proposing the aspect-of relation and exploiting syntactic patterns to handle both aspect-evaluation extraction and aspect-of extraction, the generality of the model proposed by Kobayashi et al. (2007) was enhanced.

Meng and Wang (2009) introduced a system that used the hierarchical relations among product features, which were the targets of sentiments, and units of measurement as the clues for extracting those features from reviews. Units of measurement, such as "cm" for height and width, were especially useful when feature names were not explicitly supplied in the text. Their view point of applying units of measurement to identify features was reasonable and convincing.

### 2.4 Summarization of Sentiments

Turney (2002), Tsou et al. (2005) and Meng and Wang (2009) have presented the summarization of sentiments in different ways.

Turney (2002) provided the summary of sentiments of reviews by only giving a binary result. The result indicated the average semantic orientation of subjective phrases. It was either "Thumbs Up", which was positive, or "Thumbs Down", which was negative. His approach could thus only provide a general summary of sentiments on the reviewed object as a whole but not specifically on

any individual aspects.

Tsou et al. (2005) presented the calculations of polarity of news reports numerically as the results, but this kind of presentation was less readable.

Meng and Wang (2009) generated the summary of sentiments for product reviews without computing the polarity of users' opinions on product features. Instead, they used concrete descriptions to describe a feature, for example, "small" and "thin" were used to describe the feature "size". Descriptive words which co-occurred with the features identified in reviews were used. They suggested that this kind of description provided a clearer picture than "positive" and "negative" polarity. Comparing with previous approaches (Turney, 2002; Tsou et al., 2005), the readability of this kind of presentation was relatively higher.

## 2.5 Other Relevant Issues

Apart from the studies of the above tasks, other issues related to sentiment analysis have also been addressed by different scholars.

Wilson and Wiebe (2003) presented a detailed scheme for annotating opinions and private states, such as mental and emotional states, expressed in news and other kinds of discourse. Both implicit and explicit expressions were taken into accounts. Details such as the strength, polarity and nested sources of the sentiments were included in the scheme. The scheme was expected to improve natural language processing tasks, such as information extraction and summarization.

Liu et al. (2007) addressed the issue of filtering out the low-quality product reviews. They distinguished the low-quality reviews from high-quality ones by examining the informativeness, subjectiveness and readability of product reviews. By applying this technique to filter out the low-quality reviews, the summarization of sentiments of reviews should be more convincing and reliable.

Murakami et al. (2009) proposed a scheme for annotating fundamental units of sentences, which were referred as statements. The scheme organized opinions expressed in statements into three groups, namely AGREEMENT, CONFLICT and EVIDENCE. Statements in AGREEMENT contained similar opinions. Those in CONFLICT were in contradiction. Statements in EVIDENCE supported opinions expressed in other statements. The scheme provided users with a clearer picture of how statements were related and helped users analyze the usefulness of information including sentiments on the Internet.

## 2.6 Approaches of the Current Project

Many scholars have investigated various tasks of sentiment analysis on product reviews (Liu et al., 2007; Meng and Wang, 2009; Min and Park, 2009; Popescu and Etzioni, 2005). For the current project, a sub-genre of product reviews is selected for investigation. That is video game reviews. Instead of analyzing the overall quality of a game, the project focuses on analyzing the quality of four aspects of video games.

### 2.6.1 Identification of Subjective Linguistic Units and Targets of Sentiments

Since the project focuses on a specific domain: video game reviews, domain-dependent dictionaries are prepared for the tasks of identification of subjective linguistic units and identification of targets of sentiments. Dictionaries of keywords of aspects, opinion words and intensifiers are manually compiled to be used by the main program of the system to identify subjective units (opinion words and intensifiers) and targets of sentiments (aspects of a game) in the input reviews.

### 2.6.2 Evaluation of the Polarity of Sentiments

For the task of evaluation of the polarity of sentiments, the main program first extracts intensity values with polarity of the identified subjective units from dictionaries. Next, it computes the intensity value and polarity by multiplying values of units which relate to each other. The finer-grained scale of the intensity value adopted in the system facilitates relatively precise computation than the approach of assigning binary semantic orientation to subjective units proposed by Hatzivassiloglou and McKeown (1997).

### 2.6.3 Summarization of Sentiments

Instead of only providing an overall score or comment like the approaches of Turney (2002) and Tsou et al. (2005), the main program sums up the values for each aspect and presents the relative quality of the aspects of the game to users.

# 3 System Description

The whole system comprises several components: a dictionary of intensifiers (intDict), a dictionary of keywords of aspects (keyDict), a dictionary of opinion words (opDict), input texts and the main program, sentAn, compiled from a JAVA program. In this section, the mechanism of the system is explained. The preparation of data and

resources and the implementation of the main program are reviewed afterwards.

### 3.1 Assumptions on Input Reviews

It is assumed that the input reviews are pre-processed like this: First, it is well segmented into word tokens. Second, all punctuations and emoticons are removed.

### 3.2 Identification of Subjective Linguistic Units and Targets of Sentiments

After inputting a game review into the main program, the program processes the sentences one by one. The program searches through the whole sentence and matches each token with the entries in the dictionaries. If a token is found in one of the dictionaries, the program successfully identifies the token as an opinion word, an intensifier or a keyword of aspect. For example, if the following sentence is input, the tokens existing in the sample dictionaries will be identified.

Input: 這 遊戲 的 系統 及 畫面 真的 十分 出色 .
"The system of this game and images inside are really pretty outstanding."

Entries in the dictionaries:

keyDict: {畫面, 系統, 戰鬥, 結局, ...}
"images, system, battle, ending, ..."

intDict: {真的, 十分, 很, 超, ...}
"really, pretty, very, extremely, ..."

opDict: {出色, 好, 很棒, ...}
"outstanding, good, excellent, ..."

Tokens identified:

keywords: 畫面, 系統
"images, system"

intensifiers: 真的, 十分
"really, pretty"

opinion words: 出色
"outstanding"

It should be noted that the identification of holders of sentiments is not performed, since the holder is usually the author of a game review.

### 3.3 Evaluation of the Polarity of Sentiments

First of all, in the above stage of identification, the intensity values with polarity of the intensifiers and opinion words are extracted from the dictionaries. Also, the representing aspects of the keywords are extracted. Next, the program selects the nearest opinion word for each intensifier. Hence, an opinion word is linked with one or more intensifiers. For the example in section 3.2, "出色" is linked with "真的" and "十分". Next, the value of an opinion word is multiplied by the values of its linked intensifiers. If the value of "出色"is '1.5' and the values of "真的" and "十分" are '1.5' and '2' respectively, the value of "出色" will become '4.5'. The judgement on the values is explained in section 4.2.5 below.

After that, the program selects the nearest opinion word for each keyword. The selected opinion word must be compatible with the aspect represented by the keyword. The aspect represented by a keyword is stored in keyDict together with the keyword. The compatible aspects of an opinion word are those aspects of video games on which the opinion word can be used to comment. For example, "不錯聽" (good to listen) can only be used to comment on Sound, so that it only has Sound as its compatible aspect. The compatible aspects are stored in opDict together with the opinion word. In section 4.2.5, the determination of compatible aspects of opinion words and aspects represented by keywords is further elaborated. For the selection of the opinion word for a keyword, the program extracts compatible aspects from opDict to match with the aspect represented by the keyword stored in keyDict first. Next, if they are matched, the opinion word is assigned to the keyword. For the example above, if "出色" is compatible with the aspects Image and System represented by "畫面" and "系統" respectively, "出色" will be assigned to the two keywords. After that, the value of the opinion word is added to the overall sentiment scores of the corresponding aspects. For the example, the scores of Image and System are both increased by '4.5'.

### 3.4 Summarization of Sentiments

The output of the program is expected to provide users a summary of relative quality of the four aspects of a game: System, Story, Image and Sound according to the descending order of the overall score with polarity gained by each aspect. If any of the four aspects of the game are not commented by the author, the program will automatically drop them from the summary. For the above example, the output is:

Two aspects of the game are commented by the author.
In descending order of quality, they are: System = Image
N.B.:

The Equality sign (=) indicates that the preceding aspect is of the same quality as the succeeding aspect.

The Greater-than sign (>) indicates that the preceding aspect is of higher quality than the succeeding aspect.

## 4 Data and Resources

In this section, the preparation of data and resources is reviewed. The task involved the collection of sample reviews for fine-tuning and testing the system and extraction of words for compiling the dictionaries.

### 4.1 Collection of Sample Reviews

All the texts used in this project were retrieved from a Taiwan website of video games called Bahamut[2] (Oneup Network Corporation, n.d.). Manual search for gamers' reviews on video games was performed. Three hundred reviews were collected. The collection comprised reviews on video games from 5 genres: Role-playing Game (RPG), Action Game (ACT), Racing Game (RAC), First-person Shooting Game (FPS) and Strategy Game (STG). The distribution of genres in the collection was even. There were 60 reviews from each genre. Among those 60 reviews, 52 reviews were solely used for extracting words for compiling the dictionaries. The remaining 8 reviews were manually segmented into word tokens. Six of them were used for system fine-tuning; two of them were used for system testing. Each review was mainly in Taiwan informal Chinese and was saved in an independent plain text file in UTF-8 encoding and named in the format [<filename>(s)(t).txt]. The optional character 's' was attached to the <filename> if the review was segmented into tokens. The optional character 't' was attached if the review was used for system testing. All segmented reviews were manually pre-processed. First, boundaries of sentences were identified by contextual clues like line breaking and punctuations. Next, punctuations and emoticons were removed from the article. After that, a full stop was added to each sentence boundary and space characters were inserted to mark the boundaries between word tokens. Correction of obviously mistyped words was also performed during the process. In addition, some reviews contained some special terms in English. Since those English words did not have great influence on the system, they were also segmented into tokens like Chinese words but they

[2] http://www.gamer.com.tw/index2.php

were not further processed.

## 4.2 Collection of Words for Compiling Dictionaries

For the extraction of words for compiling the dictionaries, the following procedures were performed. First, manual search for intensifiers, opinion words and keywords of aspects was done on the 260 non-segmented reviews. Next, each identified word and its attributes were stored in the corresponding dictionary in specific formats. The total numbers of entries in the dictionaries are as follows: 156 in opDict, 54 in keyDict and 33 in intDict.

### 4.2.1 Format for Storing Intensifiers

For intensifiers, they were stored in the format [<word>/<intensity_value_with_polarity>]. The numerical intensity value with polarity sign (none for positive, '-' for negative) was stored as the attribute of the word. A forward slash was used to separate the word and its attribute in the entry. For example, since the intensifier "真的" (really) had a positive polarity and its intensity value was '1.5', it was stored as [真的/1.5].

### 4.2.2 Format for Storing Opinion Words

For opinion words, they were stored in the format [<word>/<intensity_value_with_polarity>/<aspect_one>/<aspect_two>/...]. The numerical intensity value with polarity sign and the compatible aspects were stored as the attributes of the word. Forward slashes were used to separate the word and each of its attributes in the entry. For example, the opinion word "出色" (outstanding) had a positive polarity and an intensity value of '1.5', and it was compatible with all four aspects. Thus, the word and its attributes were stored as [出色/1.5/System/Story/Image/Sound].

### 4.2.3 Format for Storing Keywords of Aspects

For keywords of aspects, they were stored in the format of [<word>/<aspect>]. The corresponding aspect was stored as the attribute of the word. A forward slash was used to separate the word and its attribute in the entry. For example, since the keyword "畫面" (images) represented Image, it was stored as [畫面/Image].

### 4.2.4 Extraction of Target Words from Reviews

The identification of the words for extraction was manually done by the investigator of this project. Identified words must have at least three

occurrences in the use for expressing sentiments in the selected reviews. Unlike English, Chinese words have no clear boundaries among syntactic categories in many cases. Therefore, the part of speech of the extracted words was not restricted. To maintain the simplicity of the system, each word could exclusively appear in one particular dictionary. This was important, especially for those words which had diverse context-dependent usages. For example, the word “太少” (too few/little) could be used to modify “進步” (improve) in “進步太少” (improve too little) as an intensifier or it could be used to modify “武器種類” (weapon categories) in “武器種類太少” (too few weapon categories) as an opinion word. Thus, the usage with higher frequency was selected for these ambiguous words. Furthermore, some opinion words were stored in a combined form of the word and its intensifier in the opinion words dictionary. For example, “很棒” (excellent) was treated as a whole and stored in the opinion word dictionary. This was because “棒” (good) seldom appeared in texts without its modifier “很” (very). If it was the case that it appeared independently, it usually had the meaning of “stick” instead of “good”. Therefore, these kinds of words were stored as a whole in order to avoid ambiguity. Other phrases like “很美麗” (very beautiful) were separated into the intensifier “很” and the opinion word “美麗” (beautiful), since “很” could also modify many other opinion words. In such cases, the productivity of intensifiers was exploited to maintain a neat and systematic organization of the data.

### 4.2.5 Judgement on Attributes of Identified Words

The judgement on the attributes of the extracted words, such as the intensity value and polarity of intensifiers and opinion words, the compatible aspects of opinion words and the corresponding aspects of keywords, was also done by the investigator of the project in the following way. For determining the intensity value and polarity of intensifiers and opinion words, an eight-level scale with the interval of 0.5 between each level (2, 1.5 ... -1.5, -2 excluding 0) was applied. The value ‘2’ indicated the positive polarity of a word and it was semantically stronger than ‘1.5’, ‘1’, etc. ‘-2’ indicated the same intensity as ‘2’ but with negative polarity. The reason for assigning polarity to intensifiers was that some intensifiers like “不” has the function of negating their complements. For example, “不好” has a negative polarity while “好” has a positive polarity. Therefore, polarity was not only assigned to opinion words but also intensifiers. Assignment of values depended on the sense of the investigator towards the words. For determining the compatible

aspect of opinion words, whenever there was evidence of use of the word in commenting particular aspects, certain aspects were included as the compatible aspects. For determining the aspect represented by keywords, only one aspect was assigned to each keyword. If a keyword could refer to more than one aspect, the most frequent aspect was selected. For example, “場景” (scenes) could either refer to Story or Image. However, since it referred to Story more frequently, Story was assigned as its corresponding aspect.

## 5 Program Implementation

The algorithm of the main program is illustrated below with an example.

### 5.1 Input Handling

First, there are four variables prepared to store the sentiment scores, which are the cumulative values of the intensity values of related opinion words, for the four aspects: System, Story, Image and Sound, namely sysScr, sndScr, storyScr and imgScr respectively. Next, a piece of segmented text is read into the program. Here, the following virtual sample input is used for illustration:

> 除了 音效 出色 之外 我 還 覺得 操作 及 畫面 真的 進步 許多 . 可是 玩 到 最後 結局 令 我 很 失望 .
> “In addition to the outstanding sound effects, I think that controls and images have really improved a lot. However, I am very disappointed with the ending.”

The input is then split into sentences by detecting full stops as sentence boundaries and stored in a String array named snt. For the sample, it is divided into two:

> snt[0]: 除了 音效 出色 之外 我 還 覺得 操作 及 畫面 真的 進步 許多
> snt[1]: 可是 玩 到 最後 結局 令 我 很 失望

For each sentence, a series of procedures is done for computing the intermediate results. Word tokens in a sentence separated by space characters are further split and stored in a String array named toks. Taking the first sentence stored in snt[0] above as an example:

> toks[0]: 除了; toks[1]: 音效 ... toks[11]: 進步; toks[12]: 許多

## 5.2 Identification of Subjective Linguistic Units and Targets of Sentiments

Next, three Hashtables (opTable, intenTable and keyTable) and three Vectors (opInd, intenInd, keyInd) are created for storing information of opinion words, intensifiers and keywords identified in the sentence. By checking the whole sentence once, the index of each opinion word in toks is stored as a key in opTable and an element in opInd simultaneously. The intensity value with polarity of the word retrieved from opDict is stored as the element in opTable of the corresponding key. For intensifiers, the case is similar to that of opinion words. The indices are stored in intenTable and intenInd accordingly and the intensity values are retrieved from intDict. For keywords of aspects, keyTable stores the indices of the keywords in toks as its keys and the corresponding aspects of the keywords retrieved from keyDict as its elements. Assuming “音效” (sound effects), “操作” (controls) and “畫面” (images) are valid entries in keyDict representing Sound, System and Image respectively, “真的” (really) and “許多” (a lot) are valid entries in intDict containing the intensity values ‘1.5’ and ‘2’ respectively and “出色” (outstanding) and “進步” (improved) are valid entries in opDict containing the intensity values ‘1.5’ and ‘1’ respectively, the information of the above sentence is stored as:

opTable<key, element>: <2, 1.5>; <11, 1>
opInd[0]: 2; opInd[1]: 11

keyTable<key, element>: <1, Sound>; <7, System>;
<9, Image>
keyInd[0]: 1; keyInd[1]: 7; keyInd[2]: 9

intenTable<key, element>: <10, 1.5>; <12, 2>
intenInd[0]: 10; intenInd[1]: 12

## 5.3 Evaluation of the Polarity of Sentiments

### 5.3.1 Calculation of Intensity Values with Polarity of Opinion Words

Next, the intensity value of each opinion word stored in opTable is multiplied by the intensity values of related intensifiers stored in intenTable. The product is written as the element in opTable and the original value of that particular element is covered. This is performed based on the proximity between the opinion word and intensifiers. Each intensifier selects the nearest

opinion word by comparing the numerical differences between indices of opinion words and intensifiers stored in opInd and intenInd respectively. As a result, the relationship of opinion words to intensifiers is a one-to-many relationship. Thus, the resulting opTable for the above example becomes:

opTable<key, element>: <2, 1.5>; <11, 3>

This shows that “真的” and “許多” have selected “進步” instead of “出色” in the sentence.

### 5.3.2 Calculation of Overall Sentiment Scores of Aspects

After that, each keyword selects the nearest opinion word by comparing the numerical differences between indices of keywords and opinion words stored in keyInd and opInd respectively. The selection is under the constraint that the set of compatible aspects of the selected opinion word retrieved from opDict must include the corresponding aspect of the keyword. As a result, the relationship of keywords to opinion words is a many-to-one relationship. For the example, it is assumed that “出色” is compatible with Sound and “進步” is compatible with both System and Image. The linkages between the keywords and the opinion words are established as follows:

操作: 進步(3)
畫面: 進步(3)
音效: 出色(1.5)

The numbers inside parentheses above indicate the current intensity values of the opinion words after multiplying by the intensity values of related intensifiers, if applicable. Whenever the linkage between a keyword and an opinion word is successfully established, the intensity value of the opinion word is added to the sentiment score of the aspect represented by the keyword. Therefore, the variables storing the sentiment scores in the example have the following values up to this stage:

sysScr: 3
sndScr: 1.5
storyScr: 0
imgScr: 3

After the first sentence stored in snt[0] is processed, the subsequent sentences are processed one by one. Now, for the sake of simplicity, it is assumed that storyScr in the example has gained a value of '-2' after processing snt[1]. Then, the final sentiment scores for the sample input are:

sysScr: 3
sndScr: 1.5
storyScr: -2
imgScr: 3

### 5.4 Summarization of Sentiments

A summary of sentiments is finally generated by the system according to the different clues from the results, such as the polarity and relative intensity of the sentiment score of each aspect. The commented aspects are listed in the descending order of their sentiment scores which reflect their quality. However, the scores are used to construct the summary only. In fact, they are not explicitly shown in the summary. Furthermore, there is a Boolean variable for each aspect which indicates whether an aspect has been commented by the author or not. Thus, for the case that the score of an aspect is '0', the system can determine whether the '0' is the default value assigned in the initialization stage or the cumulative score has finally reached '0'. If an aspect is not commented, it will be dropped from the summary. For the sample review above, the output is:

Four aspects of the game are commented by the author.
In descending order of quality, they are: System = Image > Sound > Story
N.B.:
The Equality sign (=) indicates that the preceding aspect is of the same quality as the succeeding aspect.
The Greater-than sign (>) indicates that the preceding aspect is of higher quality than the succeeding aspect.

If only one aspect is commented by the author, the output will show whether its quality is high, medium or low. If the score of the aspect is positive, its quality is determined as high. If the score is '0', its quality is determined as medium. If the score is negative, its quality is determined as low. For example, if only Sound is commented in an input and its score is positive, the output is:

Only Sound of the game is commented by the author. Its quality is high.

## 6 System Results and Discussion

There were a total number of 10 reviews used for testing the system. There were several procedures done for the test. First, a gold-standard output for each test review was prepared. Next, system outputs were compared with the gold-standard outputs. The results were evaluated by accuracy A defined below:

A = (number of system outputs which match corresponding gold-standard outputs) / (total number of system outputs)

### 6.1 Preparation of Gold-standard Outputs

Three judges who were undergraduates majoring in Chinese Language and native Cantonese speakers were invited to produce the gold-standard outputs. They had been playing video games and browsing Bahamut (Oneup Network Corporation, (n.d.)) for more than 7 years.

First of all, each of them was provided with the original 10 reviews which had not been pre-processed. Next, each of the judges was asked to fill in a comment sheet in Chinese as shown in the Appendix for each review. Explanation for the comment sheet was verbally given to them. On the comment sheet, they were asked to answer the following questions based on their own judgement without referring to the system dictionaries and access to the program:

a. How many aspects (System, Story, Image and Sound) of the game do you think the author of the review has commented on?
b. List the commented aspects in descending order of their quality according to the author's comments.
c. List all the intensifiers, opinion words and keywords of aspects in the order of occurrence and provide the values of their attributes according to your own judgement. Draw a line between each pair of related words.

The answers of questions (a) and (b) formed the gold-standard output. The answers of question (c) were collected for further evaluation of the system. It was decided if any disagreement among the judges was found on the answers of questions (a) and (b) for a review, the review would be abandoned and substituted

by a new one from the Web. However, no disagreement was observed. Thus, all gold-standard outputs were used.

## 6.2 Test Result

The result of accuracy was 0.6 (6/10). There were four reviews wrongly analyzed by the system. The system outputs of two of them contradicted the gold-standard outputs in the numbers of commented aspects. All system outputs of the four reviews contradicted the gold-standard outputs in the order of the quality of aspects.

## 6.3 Discussion

Obviously, there is plenty of room for improvement of the system. After investigating the answers of question (c) on the comment sheets, several defects of the system are addressed below.

### 6.3.1 Failure in Identification of Implicit Opinions

In the test reviews, there are some opinions expressed implicitly. For example, "動畫 也 真的 像 看 電影" (watching the computer animation is really like watching a movie) expresses a positive opinion on Image by treating the computer animation as an analogy to a movie but the system fails to identify the opinion. In fact, two of the judges thought that "電影" (movie) was an opinion word and labelled it with intensity and polarity. The other judge thought that it was an opinion word but he did not provide the values of its attributes. Instead, he remarked that the whole phrase contributed a positive value to Image. Another example is "有 岔路 也 只是 拿 個 東西" (There is only something to be collected beyond the crossroads.). This example reflects the thought of the author that the map in the game bores him. Thus, it is a negative opinion on System. Similarly, the system cannot identify that opinion. However, even though these kinds of implicit expressions are clues to sentiment identification, it is difficult to have the system consider them in the identification task. This is because sentiments are expressed by the whole expressions instead of words. It is unreasonable to include them in any of the dictionaries of the system.

### 6.3.2 Insufficient Information in Dictionaries

There are some keywords and opinion words which appear in the reviews but were not included in the dictionaries. For example, "推進模式" (driving mode) is a keyword of System which appears in a review. However, it

is not included in the dictionary. Also, for some opinion words, the compatible aspects are incompletely assigned in the dictionary. For example, “細膩” (exquisite) is used to describe “配樂” (background music) of Sound. However, only Image is its compatible aspect listed in the dictionary. These defects of the dictionaries prohibit the program from identifying the mentioned aspect and the expressed sentiment. The expansion of the dictionaries is needed.

### 6.3.3 Discrepancy on Intensity Values

It is observed that intensity values of some opinion words and intensifiers marked in the dictionaries are different from those provided by the judges. For example, “喜歡” (like) has the value of ‘1’ in the dictionary while two of the judges thought that it should have the value of ‘1.5’. This situation indiscriminately occurs in all ten reviews. The occurrence of this discrepancy is unavoidable because the dictionaries were compiled without the assistance of other ready-made sentiment dictionaries. Therefore, subjective judgement on words’ attributes provokes the discrepancy.

### 6.3.4 Algorithm-driven Problems

First, relationships between keywords and opinion words are established by computing the proximity of the two. A keyword selects the nearest opinion word which is compatible with its corresponding aspect. However, this method does not always correctly establish the relationships. For example, in a sentence like “畫面 真的 不錯 但 音樂 卻 不是 很 好” (images are really not bad but the music is not very good), since there is only one token between “不錯” (not bad) and “音樂” (music) in the sentence but there are three tokens between “音樂” and “好” (good), the program selects the nearer opinion word “不錯” for “音樂”. However, “不錯” is actually used to comment on “畫面” (images) while the actual opinion word which describes “音樂” is “好”. Hence, it shows that proximity is an insufficient factor in determining the assignment of opinion words to keywords.

Second, the algorithm works on a sentence-by-sentence basis. If opinions are not expressed in the same sentence of relevant keywords, the system will fail to identify the opinions. For example, the system fails to identify “單調” (dull) as an opinion towards “地圖” (map) in “但 就是 遊戲 的 地圖 沒有 像 迷宮 一樣 . 有 岔路 也 只是 拿 個 東西 . 有 一點 太 單調 沒有 挑戰性”. (The map in the game does not look like a maze. There is only something to be collected beyond the crossroads. It is a bit dull and not challenging.) This is because “地圖” and “單調” are in different

sentences. To solve problems of this kind, in addition to sentence-level processing, discourse-level processing is needed.

Third, adopting one-to-many relationships for opinion words to keywords facilitate the system to correctly establish the relationships between the pairs such as “操作” (controls) and “進步” (improved), and “畫面” (images) and “進步” in “操作 及 畫面 真的 進步 許多” (controls and images have really improved a lot). However, the system fails to handle the cases of linking more than one opinion word to a keyword. For example, there are actually two opinion words “精美” (elegant) and “流暢” (smooth) describing “畫面” in “畫面 真是 精美 又 流暢” (images are really elegant and smooth). However, the system only selects the closer one for “畫面”, which is “精美”. This is because the algorithm can assign only one opinion word to each keyword. It fails to detect that there is another opinion word describing the keyword. To solve problems of this kind, conjunctions like “又” (and) or other connectives can be used as clues to identify opinion words which all describe the same keyword.

## 7 Future Work

The work so far is just the foundation stone of the development of the system. More efforts on the future work are expected. Firstly, dictionaries should be expanded to include sufficient entries. Secondly, the values of attributes in the dictionaries should be revised with the assistance of other resources such as corpora with sentiment annotations. Thirdly, for assigning opinion words to keywords, other contextual clues like connectives should be exploited in addition to proximity and discourse-level processing should be done in addition to sentence-level processing. Furthermore, methods proposed by different scholars can be adopted to improve the performance on the tasks of subjective units and targets of sentiments identification and polarity evaluation. Last but not least, apart from showing the relative quality of various aspects in the summary, attempts to generate individual scores of aspects in a fixed scale should be made. For example, a five-star grading method can be developed to summarize a review on a game which has good Image, satisfactory Sound and poor System and Story in the following format:

Image: five stars
Sound: three stars
System: one star
Story: one star

## 8 Conclusion

This project developed a system for analyzing sentiments expressed in unannotated segmented video game reviews in Chinese. The system aims at handling major tasks of sentiment analysis: identification of subjective linguistic units and targets of sentiments, evaluation of the polarity of sentiments and summarization of sentiments. The identification tasks are performed with the assistance of the domain-dependent dictionaries. In addition to the dictionaries, the evaluation of polarity depends heavily on the proximity between intensifiers and opinion words and that between opinion words and keywords of aspects. The summarization of sentiments is presented in a short summary of the relative quality of aspects (System, Story, Image and Sound) of video games. Throughout the development of the system, the steps of collecting data, compiling dictionaries, implementing the main program and testing and evaluating the system were gone through sequentially. The system testing yielded fair results. More efforts should be made to improve the system in the future.

**References**


Fillmore, C. J., & Baker, C. F. (n.d.). *FrameNet*. Retrieved from http://framenet.icsi.berkeley.edu/

Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the Semantic Orientation of Adjectives. *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, pp. 174-181. Madrid, Spain.

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*, pp. 1-8. Sydney, Australia.

Kobayashi, N., Inui, K., & Matsumoto, Y. (2007). Extracting Aspect-Evaluation and Aspect-of Relations in Opinion Mining. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 1065–1074. Prague, Czech Republic.

Liu, J., Cao, Y., Lin, C. Y., Huang, Y., & Zhou, M. (2007). Low-Quality Product Review Detection in Opinion Summarization. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 334-342. Prague, Czech Republic.

Meng, X., & Wang, H. (2009). Mining User Reviews: from Specification to Summarization. *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pp. 177-180. Suntec, Singapore.

Min, H. J., & Park, J. C. (2009). Toward finer-grained sentiment identification in product reviews through linguistic and ontological analyses. *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pp. 169-172. Suntec, Singapore.

Murakami, K., Masuda, S., Matsuyoshi, S., Nichols, E., Inui, K., & Matsumoto, Y. (2009). Annotating Semantic Relations Combining Facts and Opinions. *Proceedings of the Third Linguistic Annotation Workshop*, pp. 150-153. Suntec, Singapore.

Oneup Network Corporation (n.d.). *Bahamut*. Retrieved from http://www.gamer.com.tw/index2.php

Pang, B., & Lee, L. (2004). A sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. *Proceedings of the 42nd Meeting of the Association for Computational Linguistics*, pp. 271-278. Barcelona, Spain.

Popescu, A. M., & Etzioni, O. (2005). Extracting Product Features and Opinions from Reviews. *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pp. 339-346. Vancouver, Canada.

Riloff, E., Wiebe, J., & Wilson, T. (2003). Learning Subjective Nouns using Extraction Pattern Bootstrapping. *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pp. 25-32. Edmonton, Canada.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., & Scheffczyk, J. (2006). *FrameNet II: Extended Theory and Practice*. Retrieved from http://framenet.icsi.berkeley.edu/index.php?option=com_wrapper&Itemid=126

Tsou, B. K. T., Kwong, O. Y., Wong, W. L., & Lai, T. B. Y. (2005). Sentiment and Content Analysis of Chinese News Coverage. *International Journal of Computer Processing of Oriental Languages, 18*(2), 171-183.

Turney, P. D. (2002). Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 417-424. Philadelphia, Pennsylvania, USA.

Wiebe, J. M., Bruce, R. F., & O'Hara, T. P. (1999). Development and Use of a Gold-Standard Data Set for Subjectivity Classifications. *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pp. 246-253. College Park, Maryland, USA.

Wilson, T., & Wiebe, J. (2003). Annotating Opinions in the World Press. *Proceedings of the 4th ACL SIGdial Workshop on Discourse and Dialogue*, pp. 13-22. Sapporo, Japan.

**Appendix**

Sample Comment Sheet

遊戲評論編號: _____

分析員: _____

日期: _____

1. 你認為這篇評論的作者對遊戲的哪個方面(System, Story, Image and Sound)提出了意見？

   _____

2. 請根據作者的意見，由高至低排列出遊戲各方面的質素。

   _____

3. 請依其出現之先後次序，列出所有加強詞、意見詞及遊戲各方面的關鍵詞，並寫上其屬性值。最後於相互對應的詞語之間畫一條線。