

國立臺灣大學電機資訊學院資訊工程學研究所



博士論文

Graduate Institute of Computer Science and Information Engineering

College of Electrical Engineering and Computer Science

National Taiwan University

Doctoral Dissertation

意在言外？微文本中情緒、合法性與反諷之辨識與分析

Beyond Literal Meanings: Recognition and Analysis  
of Emotions, Legality and Irony in Microtexts

唐宜傑

Yi-jie Tang

指導教授：陳信希 博士

Advisor: Hsin-Hsi Chen, Ph.D.

中華民國 104 年 6 月

June 2015



國立臺灣大學博士學位論文  
口試委員會審定書

意在言外？微文本中情緒、合法性與反諷之辨識與分析

Beyond Literal Meanings: Recognition and Analysis of  
Emotions, Legality and Irony in Microtexts

本論文係唐宜傑君（學號D95922018）在國立臺灣大學資訊工程學系完成之博士學位論文，於民國 104 年 6 月 30 日承下列考試委員審查通過及口試及格，特此證明

口試委員：

陳信芬

（指導教授）

鄭上巳

張嘉惠

古倫維

宋麗梅

陳志華

趙坤茂

系主任

## 誌謝



本論文得以順利完成，要特別感謝指導教授陳信希老師。由於他的指導，使我能從語言學領域轉換到截然不同的資訊工程學研究環境，並得以將研究成果發表在數個與自然語言處理相關的國際性研討會上，最後還能完成這份博士論文，這些都是我在進入資訊所之前所無法想像的。

本論文承蒙陳光華老師、鄭壬壬老師、古倫維老師、張嘉惠老師在實驗設計和演算法方面給予指教，而宋麗梅老師與蘇以文老師也在語言學方面給予指正和提醒。他們確實使我發現自己的不足之處，在此也向他們表達感謝。

另外也要謝謝實驗室碩士班同學李長曄，他收集的微網誌語料庫對於論文的各項實驗提供了不少助益。

## 摘要



傳統自然語言處理所偵測和辨識的目標，多是可與表面形式直接連結的元素，然而語言中亦存在著非字面或非詞彙的層面，這些現象並無法直接由字面解析的方式來了解。而在微文本（microtext）中，由於文字篇幅受到限制，因此這些方面的分析變得更加困難。不論在情感分析、意見探勘、問答系統、或對話系統等應用中，這些問題皆可能而造成障礙。在本研究中，我們探討線上溝通中三個超越字面層面的現象：情緒、合法性、反諷，並以微網誌平台上的訊息和政府公布的短語為語料來進行偵測演算法研究和相關語言分析。

在微網誌的情緒研究方面，我們採用微網誌作者常用的圖形化表情符號為情緒標記，來建立正面情緒和反面情緒微網誌資料集。情緒的偵測採用分類演算法進行，除了文字特徵之外，還加入社交關係、使用者行為、相關度等因素作為特徵。研究發現，若適當搭配文字特徵和特定非文字特徵，可達到最佳偵測結果。此外，我們也探討貼文者與回應者之間的情緒轉換，並就詞頻、語意、情感等方面來比較微網誌文字內容和線上長篇文章間的差異。

隨著近來線上行銷活動持續成長，大量與行銷相關的微文本內容亦在線上產生。這些文本可能包含不應讓使用者列入參考的不當資訊，但這種誤導性質經常無法簡單地從訊息本身看出。不論對於網站讀者、廣告主、廣告服務商和政府管理單位來說，如何辨識這些不當行銷資訊，都已成為一項重要的課題。本研究以政府公布的違規廣告敘述和購物網站商品描述作為非法與合法廣告資料集，透過單一分類和二元分類演算法進行合法性預測，並以單詞組、同義詞典、政府規範內容和相對頻率比率對數等特徵進行實驗。結果發現結合單詞組和相對頻率比率對數作為特徵，可得到最佳的結果。相對頻率比率對數也用於對非法廣告資料集進行動詞組的探勘，這些動詞組皆由動詞與受詞組成，所形成的非法廣告用詞表可讓廣告主和政府單位作為辨識廣告合法性的參考依據。此外，本研究也實作一套不當線上廣告辨識系統，希望能為相關機構和使用者提供自動辨識機制，以節省人力並減少此類不

當行銷活動所帶來的危害。

反諷是一種少見但具有強烈效果的表達方式。英語的 verbal irony 一詞可指字面語意和真實語意相反或有程度差異的表達方式。本研究將焦點集中在以正面字面語意來表達負面實際語意的短語，並以微網誌作為語料，進行以下研究：(1) 中文反諷語料庫的建構 (2) 反諷語言結構的探討 (3) 反諷線索的歸納 (4) 反諷成分的辨識。

為了儘可能找到足夠的反諷文字型式，本研究以表情符號作為情緒極性標記，並以 NTUSD 意見詞典和微網誌正負面情緒詞典作為情緒判斷依據，透過反復式自助法 (bootstrapping) 來尋找反諷訊息，也就是先觀察特定反諷文字形式，接著以此形式為基礎透過半自動方法找出微網誌語料中的反諷訊息，再以其中新發現的反諷文字形式重複進行原步驟，直到無法發現新形式為止。以此方法，我們成功地建立了第一個中文反諷語料庫。在反諷訊息辨識上，則採用條件隨機域 (CRF) 作為演算法，並以中文詞以及其詞類標記作為特徵來進行。此辨識方法可以減少前述方法中人工介入的程度。在對反諷結構進行分析後，我們認為有三項成分構成反諷文字：(1) 反語 (2) 情境資訊 (3) 修辭成分。這些成分亦明確標示在我們的反諷語料庫中。

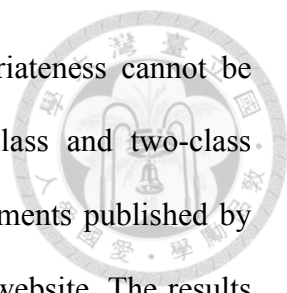
# ABSTRACT



The non-literal or non-lexical aspects of communication cannot be interpreted directly and literally. The identification and analysis of real intent beyond literal meanings is a challenging task in natural language processing, especially when working on microtexts such as microblogs that are limited to 140 characters. The recognition and analysis of these components are crucial for many applications including sentiment analysis, opinion mining, question answering and chatterbots. In this study, emotion recognition, online advertising legality identification and verbal irony analysis are examined.

In the emotion recognition experiments, the generation of user emotions on a microblogging platform is modeled from both writers' and readers' perspectives. Graphic emoticons, which are commonly used to express users' emotions, serve as emotion labels so that microtext emotion datasets can be constructed. To build classifiers for the emotion identification task, support vector machine (SVM)-based algorithms are adopted. In addition to textual features, non-verbal factors, including social relation, user behavior and relevance degree, are also used as features. The experimental results show that the combination of textual, social and behavioral features can be used to achieve the best emotion-prediction performance. The emotional transitions from the poster to the responder in a conversation are also analyzed and predicted in this study.

As online advertising continues to grow, Internet users, advertisers, online advertising platforms and the authorities all have the need to avoid or prevent the issues that false and/or misleading advertisements can potentially cause. Many of these false



advertising messages are present in short texts, and their appropriateness cannot be easily interpreted. This problem is addressed by building one-class and two-class classifiers with datasets consisting of short illegal advertising statements published by the government and product descriptions from an online shopping website. The results show that the models using the log relative frequency ratio (logRF) combined with unigrams as features achieve the best performance. The logRF values are also used to mine verb phrases that are typically used in illegal advertisements. These verb phrases can be used as a reference for both the advertisers and the authorities. A web-based false advertisement recognition system was also built in this study using the techniques applied to the above experiments in order to reduce human effort in filtering false advertising messages and help protect Internet users from misleading advertising.

In verbal irony, the literal meaning of an utterance can be the opposite of what is actually meant. For simplification, this study focuses on ironic expressions in which negative actual meanings are represented by positive words. Ironic messages in microblogs are infrequent and cannot be identified by simply examining the literal meanings of the words. To construct a Chinese irony corpus, ironic messages are collected from microblogs based on emoticon use, linguistic forms and sentiment polarity through a bootstrapping approach. Five types of irony patterns are found in the collected ironic messages. The structure of ironic expressions is also analyzed, and three types of elements are found to form an ironic expression. A conditional random field (CRF)-based approach is used to automatically identify irony elements and ironic messages and reduce the human effort in the bootstrapping approach of irony pattern discovery.

# TABLE OF CONTENTS



|                                          |     |
|------------------------------------------|-----|
| 誌謝.....                                  | i   |
| 摘要.....                                  | ii  |
| ABSTRACT.....                            | iv  |
| TABLE OF CONTENTS.....                   | vi  |
| LIST OF FIGURES.....                     | x   |
| LIST OF TABLES.....                      | xii |
| Chapter 1 Introduction.....              | 1   |
| Chapter 2 Real Intent in Microtexts..... | 4   |
| 2.1 What Is a Microtext?.....            | 4   |
| 2.2 Microblogging Platforms.....         | 6   |
| 2.2.1 Microtexts on Twitter.....         | 6   |
| 2.2.2 Microtexts on Plurk.....           | 6   |
| 2.3 Intent Hidden in Microtexts.....     | 7   |
| 2.3.1 Emotions.....                      | 8   |
| 2.3.2 Adverting Legality.....            | 10  |
| 2.3.3 Irony.....                         | 11  |
| Chapter 3 Emotions in Microblogs.....    | 13  |
| 3.1 Dataset.....                         | 14  |
| 3.1.1 Emoticons on Plurk.....            | 14  |
| 3.1.2 The Plurk Dataset.....             | 15  |
| 3.2 Reader/Writer Perspective.....       | 16  |

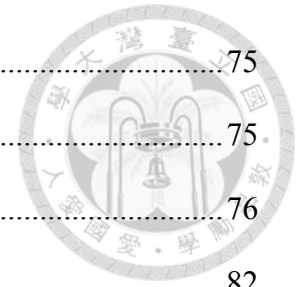


|           |                                            |    |
|-----------|--------------------------------------------|----|
| 3.2.1     | Reader Perspective .....                   | 17 |
| 3.2.2     | Writer Perspective .....                   | 17 |
| 3.2.3     | Reader and Writer Perspective .....        | 18 |
| 3.3       | Emotion Modeling.....                      | 18 |
| 3.3.1     | Text Features (T).....                     | 18 |
| 3.3.2     | Social Relation (S) .....                  | 19 |
| 3.3.3     | User Behavior (B) .....                    | 20 |
| 3.3.4     | Relevance Degree (R) .....                 | 21 |
| 3.3.5     | Normalization.....                         | 21 |
| 3.4       | Results .....                              | 22 |
| 3.4.1     | Text Features ( <i>T</i> ) .....           | 22 |
| 3.4.2     | Combination of Feature Sets .....          | 23 |
| 3.4.3     | Different Perspectives .....               | 25 |
| 3.4.4     | Writer Model.....                          | 26 |
| 3.5       | Writer-Reader Emotion Transition.....      | 26 |
| 3.5.1     | Sentiment Word Mining.....                 | 27 |
| 3.5.2     | Analysis of Mined Words .....              | 30 |
| 3.5.3     | Writer-Reader Emotion Prediction.....      | 31 |
| 3.6       | Microtexts Compared to Regular Texts ..... | 33 |
| 3.6.1     | Microblog Corpus vs. Balanced Corpus ..... | 34 |
| 3.6.2     | Analysis of Lexical Semantics .....        | 34 |
| 3.6.3     | Literal Text vs. Spoken Language.....      | 40 |
| 3.6.4     | Analysis of Sentiment Polarity.....        | 40 |
| Chapter 4 | Advertising Legality .....                 | 43 |
| 4.1       | Illegal Advertising.....                   | 43 |



|           |                                                               |    |
|-----------|---------------------------------------------------------------|----|
| 4.2       | Datasets .....                                                | 45 |
| 4.3       | Recognition Models .....                                      | 46 |
| 4.3.1     | One-Class Classifiers .....                                   | 47 |
| 4.3.2     | Binary Classifiers .....                                      | 48 |
| 4.4       | Overstated Verb Phrase Mining .....                           | 54 |
| 4.5       | Building False Advertising Recognition System .....           | 56 |
| 4.5.1     | Pre-processing Module.....                                    | 56 |
| 4.5.2     | Recognition Module.....                                       | 57 |
| 4.5.3     | Explanation Module .....                                      | 58 |
| 4.5.4     | User Interface .....                                          | 58 |
| Chapter 5 | Verbal Irony.....                                             | 61 |
| 5.1       | Irony and Sarcasm.....                                        | 61 |
| 5.2       | Irony Corpus Construction .....                               | 64 |
| 5.2.1     | Finding Irony from an Emotion-Tagged Corpus.....              | 64 |
| 5.2.2     | Candidates Extraction .....                                   | 66 |
| 5.2.3     | Pattern Mining.....                                           | 66 |
| 5.3       | Irony Patterns .....                                          | 68 |
| 5.3.1     | Degree Adverbs + Positive Adjective .....                     | 68 |
| 5.3.2     | The Use of Positive Adjective with High Intensity.....        | 69 |
| 5.3.3     | The Use of Positive Noun with High Intensity .....            | 70 |
| 5.3.4     | The Use of 很好 (very good) .....                               | 70 |
| 5.3.5     | 可以再...一點 (“It’s okay for someone/something to be worse”)..... | 71 |
| 5.4       | Linguistic Structure of Irony .....                           | 72 |
| 5.4.1     | Ironic Word/Phrase .....                                      | 72 |
| 5.4.2     | Contextual Information .....                                  | 73 |

|           |                                    |    |
|-----------|------------------------------------|----|
| 5.4.3     | Rhetoric .....                     | 75 |
| 5.4.4     | Corpus Labeling .....              | 75 |
| 5.5       | Clues to Irony .....               | 76 |
| 5.6       | Irony Recognition.....             | 82 |
| 5.6.1     | CRF-based Algorithms.....          | 82 |
| 5.6.2     | Irony Element Identification ..... | 83 |
| 5.6.3     | Ironic Message Identification..... | 85 |
| Chapter 6 | Conclusion.....                    | 88 |
| REFERENCE | .....                              | 91 |

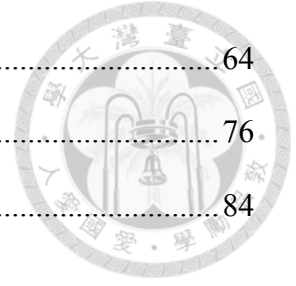


# LIST OF FIGURES



|            |                                                                                                                              |    |
|------------|------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 2.1 | A conversation on Plurk .....                                                                                                | 6  |
| Figure 3.1 | Emoticons as positive and negative labels.....                                                                               | 14 |
| Figure 3.2 | Different emotion generation models on Plurk.....                                                                            | 16 |
| Figure 3.3 | Category distribution of emotion words .....                                                                                 | 31 |
| Figure 3.4 | Sense distribution in terms of 12 large categories of the top<br>100 words.....                                              | 38 |
| Figure 3.5 | Sense distribution in terms of 12 large categories of the top<br>3,000 words .....                                           | 39 |
| Figure 3.6 | Sense distribution in terms of 94 middle categories of the<br>top 3,000 words .....                                          | 39 |
| Figure 3.7 | Ratio of sentiment words in top $n$ high-frequency words in the<br>Plurk and Sinica corpora .....                            | 41 |
| Figure 3.8 | Ratio of sentiment words in top $n$ high-frequency words in the<br>Plurk corpus .....                                        | 42 |
| Figure 3.9 | Ratio of sentiment words in top $n$ high-frequency words in<br>Sinica Corpus .....                                           | 42 |
| Figure 4.1 | Illegal food advertising statements from an advertisement provided<br>by the government and their English translations ..... | 44 |
| Figure 4.2 | System architecture of <b>FAdR</b> .....                                                                                     | 57 |
| Figure 4.3 | An illegal advertisement example.....                                                                                        | 59 |
| Figure 4.4 | Screenshot of the FAdR recognition results .....                                                                             | 60 |

|            |                                                  |    |
|------------|--------------------------------------------------|----|
| Figure 5.1 | The bootstrapping procedure. ....                | 64 |
| Figure 5.2 | An example message with ironic element tags..... | 76 |
| Figure 5.3 | Feature template used in CRF++ .....             | 84 |



# LIST OF TABLES



|           |                                                                                                                                     |    |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------|----|
| Table 3.1 | Accuracies of different feature sets.....                                                                                           | 22 |
| Table 3.2 | Accuracies of models with different feature combinations.....                                                                       | 24 |
| Table 3.3 | Similarity among emotion transition datasets .....                                                                                  | 27 |
| Table 3.4 | Accuracies of emotion prediction .....                                                                                              | 33 |
| Table 3.5 | Topic statistics of Sinica 3.0 Corpus.....                                                                                          | 34 |
| Table 3.6 | Taxonomy of <i>Cilin</i> .....                                                                                                      | 36 |
| Table 3.7 | List of top 10 high $\log\text{RF}_{\text{Plurk-Sinica}}$ words and top 10 high<br>$\log\text{RF}_{\text{Sinica-Plurk}}$ words..... | 38 |
| Table 3.8 | Distribution of sentiment words in NTUSD sentiment dictionary.....                                                                  | 41 |
| Table 4.1 | Accuracies of classifiers in different rejection rates .....                                                                        | 47 |
| Table 4.2 | Accuracies of advertising legality recognition models .....                                                                         | 49 |
| Table 4.3 | Illegal advertising statement examples announced by the government.....                                                             | 49 |
| Table 4.4 | Individual accuracies of illegal and legal advertising recognition<br>on food dataset.....                                          | 54 |
| Table 4.5 | Individual accuracies of illegal and legal advertising<br>recognition on cosmetic dataset .....                                     | 54 |
| Table 4.6 | Example illegal verb phrases mined from the FOOD and COS datasets.....                                                              | 55 |
| Table 5.1 | Likelihoods of irony for random 1,000 messages with specific patterns.....                                                          | 78 |
| Table 5.2 | Experimental results of irony element identification.....                                                                           | 85 |

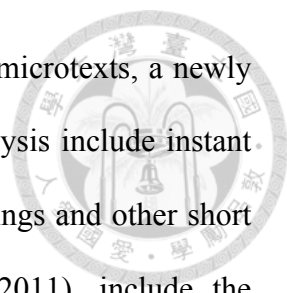
# Chapter 1 Introduction



Literal communication refers to the conveyance of ideas, sentiments and intentions through devices that express the original or usual sense of words. The research of non-literal or non-lexical expression identification has drawn researchers' attention in recent years. Katz and Giesbrecht (2006) used meaning vectors for literal and non-literal expression classification. Some previous work (e.g., Sporleder and Li, 2009) explored non-literal communications by focusing on the interpretation and recognition of idioms. Jimenez, Becerra and Gelbukh (2013) introduced a system to discriminate between literal and figurative usage of phrases. In this study, the analysis and identification of three aspects of online communication are explored beyond literal meanings: emotion, advertising legality and verbal irony.

Traditional tasks in natural language processing (NLP) include word segmentation, word sense disambiguation, grammatical structure parsing, POS tagging, machine translation, named entity recognition, etc. Most of them are performed on traditional long texts and are based on the assumption that linguistic information can be retrieved from literal meanings. The data sources of traditional texts include books, news articles, journal articles, and classical literature. Such texts consist mainly of grammatically well-structured sentences and contain few "errors." In the case of any ungrammatical sentences, typos and misprints, they are usually considered noise and thus are finally ignored in NLP tasks.

In recent years, many studies have turned their attention to more challenging tasks such as opinion mining and sentiment analysis. With the emergence of a large number of contents generated by average Internet users, newer and more informal language uses have been taken into account.



To perform the identification and analysis tasks in this study, microtexts, a newly emerging category of texts, are used. The objects of microtext analysis include instant messages, Short Message Service (SMS) messages, microblog postings and other short texts. The characteristics of a microtext, suggested by Ellen (2011), include the following:

- (1) Individual author contributions are very brief and almost always less than a paragraph.
- (2) The grammar used by the authors is generally informal.
- (3) The text is “semi-structured” by traditional NLP definitions.

The datasets used in this study were collected from a microblogging platform and short advertising statements made public by the government. As can be seen in this study, the properties specific to microtexts make the processing of non-literal and non-lexical communications in these datasets challenging.

Emotions have only an indirect link to lexical senses. The recognition of emotions expressed in microblogs must not only deal with literal meanings of words but also take non-lexical factors such as social relation and user behavior into account. In Chapter 3, emotion-modeling approaches are proposed from both writers’ and readers’ perspectives. The disagreement between the emotions of the writer and the reader is also analyzed. In addition, the differences between microblogs and regular online articles are studied in terms of sentiment, word frequency and semantics.

Online advertising is an important way of delivering promotional messages today. Misleading and overstated advertisements are serious issues for Internet users, advertisers and the authorities. In Chapter 4, online advertising legality identification



models using textual and the log relative frequency ratio as features are proposed. A web-based system for false online advertisement recognition was also constructed.

Recognizing irony is one of the most challenging issues in natural language processing. In some previous research (e.g., Davidov, 2010; Filatova, 2012), the identification of sarcasm was studied, but a clear definition of “sarcasm” was not included. In Chapter 5, we deal with only ironic expressions in which the literal meaning is the opposite of the actual meaning. A bootstrapping approach is proposed to construct a Chinese irony corpus and to discover irony patterns. The elements of irony and the clues to irony are also analyzed. Finally, a conditional random field (CRF)-based algorithm is used to detect irony elements and ironic messages in order to improve the bootstrapping procedure.

## Chapter 2 Real Intent in Microtexts




Microtexts are relatively short and provide less information than traditional articles, making NLP tasks performed on these short texts challenging. Discovering actual intent in microtexts beyond their literal meanings is even more difficult. In this chapter, the concept of microtext is introduced. The scopes of the three main topics—emotion modeling, advertising legality identification, and irony analysis, as well as related work—are covered.

### 2.1 What Is a Microtext?

In recent years, a large number of short texts have been generated by users on instant messaging platforms, online social media platforms and mobile devices. These texts, which have been referred to as microtexts in some studies (Ellen, 2011), are relatively short and informal compared to news, blog posts and other online articles.

Ellen (2011) proposed a working definition of “microtext” as follows.

- (1) Individual author contributions are very brief, consisting of as little as a single word, and almost always less than a paragraph. Frequently the contribution is a single sentence or less.
- (2) The grammar used by the authors is generally informal and unstructured, relative to the pertinent domain. Errors and abbreviations are more common.
- (3) The text is “semi-structured” by traditional NLP definitions since it contains some meta-data in proportion to some free-text. At a minimum, all microtext has a minute-level timestamp and a source attribution (author).



A microtext may be an SMS message, instant message, chatroom message, voicemail transcription, microblog post, blog comment, or query used for an intelligent personal assistant on a smartphone. In the PTT bulletin board system ([telnet://ptt.cc](http://ptt.cc)), a popular online forum in Taiwan, a large number of positive comments (推) and negative comments (噓) are also posted every day. Every PTT comment is limited to 55 characters, and thus makes another form of microtext.

Based on our observation, the characteristics of microtexts compared to those of traditional long texts are described below.

- (1) Fewer words or characters are used. Occasionally, only non-textual elements, such as emoticons and URLs, are present.
- (2) Informal words, phrases, symbols and grammar are frequently used. Typos, errors, abbreviations, slang and jargon are common.
- (3) Components of a formal article structure, such as the title, subtitle, introduction and conclusion, are generally missing.
- (4) No traditional tags can be added, but hashtags can be added anywhere within a short text.
- (5) A microtext can be part of a conversation. In this case, the contextual information can usually be obtained.

Since traditional techniques of natural language processing for long texts do not translate well to microtext (Ellen, 2011), new algorithms must be developed for microtext in order to increase the performance of NLP and data mining tasks.



## 2.2 Microblogging Platforms

### 2.2.1 Microtexts on Twitter

As of March 31, 2015, the Twitter microblogging platform had more than 302 million active users.<sup>1</sup> Short messages sent on Twitter, or “tweets,” can never exceed the 140-character limit. This is why Twitter has soon become a critical resource of short text analysis since its launch in 2006. Due to the characteristics of microtexts described in Section 2.1, performing common NLP tasks on Twitter has been regarded as a challenge. According to Kaufmann (2010), the performance of the named entity recognition task in tweets was “extremely poor.” Dent and Paul (2011) also found question detection within Twitter to be a major challenge. The traditional part-of-speech tagging algorithms were also found to perform poorly, with error rates up to 10 times higher than those on the newswire (Derczynski et al., 2013).

### 2.2.2 Microtexts on Plurk

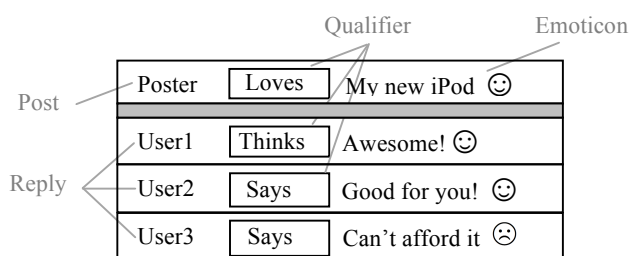


Figure 2.1 A conversation on Plurk

In addition to Twitter, Plurk has once become one of the largest online social networks in Taiwan and in a few other countries. Messages on Plurk are also limited to 140 characters. From this viewpoint, Plurk is similar to Twitter and other microblogging platforms. Unlike Twitter, however, Plurk also acts like an instant messaging system

<sup>1</sup> <https://about.twitter.com/company>

because a user can send messages to specific users and see replies as soon as they respond. A post and its replies are grouped within a box on the screen, indicating that they are messages from the same conversation. Every post can be given one of the 18 optional “qualifiers,” which are one-word verbs indicating the poster’s action or feelings: *loves, likes, shares, gives, hates, wants, wishes, needs, will, hopes, asks, has, was, wonders, feels, thinks, says, and is*.

Figure 2.1 shows a typical conversation on Plurk. In this conversation, the first line was entered by a poster. He chose "loves" as the qualifier, stating that he "loves the iPod." The following messages were entered by other users as replies to the poster. Their messages are followed by graphic emoticons that express their emotions.

In this study, the Plurk Corpus (Lee, 2010), which contains 20,265,405 Traditional Chinese posts generated from April 1 to October 31, 2009, is used for microtext processing.

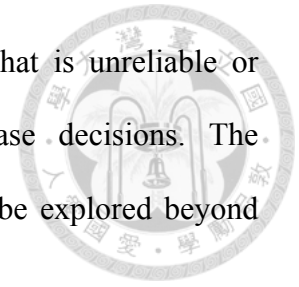
### **2.3 Intent Hidden in Microtexts**

The objective of the study is to identify and analyze several types of intent that are hidden or “disguised” in short texts and therefore cannot be accessed directly by analyzing the literal meanings or senses.

Although the emotions behind a short text can usually be linked to the literal meanings of the words that make up the text, the actual emotions to be expressed by the author can be affected by specific conditions and contexts present in a conversation. This can be one of the reasons authors choose to use emoticons in their microblog posts in addition to textual expressions. In this case, non-verbal and/or non-literal information needs to be obtained in order to identify the actual emotions.

Online advertising in violation of regulations provides false or misleading

information. The advertiser's intent may be to use information that is unreliable or should not be consulted by customers to influence their purchase decisions. The recognition of the appropriateness of such advertisements should be explored beyond the literal meanings.



Irony is another phenomenon that does not exhibit real intent in the surface form of an utterance. Irony should be precisely identified in order to correctly obtain the actual meanings behind an expression. If an ironic expression is incorrectly treated as a regular expression, the literal meaning, which is opposite to the author's real intent, will be retrieved, causing a serious problem for any natural language understanding tasks and semantic analysis.

### 2.3.1 Emotions

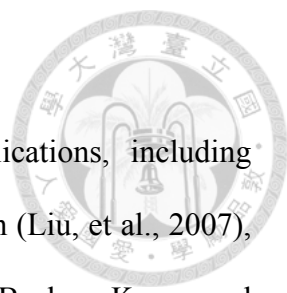
Interactions between emotion and language occur in many aspects of communication. In spoken language, the speaker's emotions can be revealed in lexical levels as well as non-lexical levels such as phonetic features, gestures and facial expressions. In written texts, especially microtexts, emotions may not be able to be decoded by solely examining the verbal form, and the features and factors beyond literal meanings thus need to be taken into account for emotion detection. For example, the following two messages posted by two different Plurk users exhibit totally different emotions, while the same words are used:

(1) 該睡了 :-) 晚安!!

(It's time for bed. :-) Good night!!)

(2) 該睡了 ==, 晚安

(It's time for bed ==. Good night.)



Emotion mining and detection are crucial for many applications, including customer care (Gupta, Gilbert, and Fabrizio, 2010), sales prediction (Liu, et al., 2007), game animation (Bernhaupt et al., 2007), and robot simulation (Becker, Kopp, and Wachsmuth, 2004). Typical tasks in emotion mining include capturing people's feelings, predicting their reactions to events, and generating suitable emotions.

Since a large number of online articles have become available, many studies have started collecting these long texts and implementing specialized algorithms to automatically identify emotions in this kind of content. Mishne (2005) adopted a corpus of LiveJournal articles that included indications of writers' moods to train mood classifiers with SVM. Mishne and Rijke (2006) used a blog corpus to identify the intensity of community mood during some given time intervals. Jung, Choi, and Myaeng (2007) also focused on the mood-classification problem in LiveJournal.

Some studies have tried to obtain lists of emotion-related words from Chinese blog articles and use classification algorithms to identify author's emotions. Yang, Lin, and Chen (2007a) used articles posted on the Yahoo Kimo Blog website as corpora to build emotion lexicons. A collocation model was proposed to learn emotional lexicons from blog articles. Emotion classification at the sentence level was experimented on by using the mined lexicons to demonstrate their usefulness. Yang, Lin, and Chen (2008) further investigated emotion classification of blog corpora using SVM and conditional random field (CRF) machine-learning techniques. The emotion classifiers were trained at the sentence level and applied to the document level. Their experiments showed that CRF classifiers outperformed SVM classifiers.

Lin, Yang and Chen (2007) pioneered reader emotion analysis with an emotion-tagged corpus generated from Yahoo Kimo news articles. They later classified

documents into reader emotion categories with SVM and Naïve Bayes classifiers (Lin, Yang and Chen, 2008). Besides classification, Lin and Chen (2008) proposed algorithms based on pairwise loss minimization (PLM) and emotional distribution regression (EDR) to rank reader emotions. They showed that EDR was better at predicting the most popular emotion, while PLM produced ranked lists that had higher correlation with the correct lists. Yang, Lin, and Chen (2009) further introduced the application of emotion analysis from both writers' and readers' perspectives. The relationships between writer and reader emotions were discussed in their work.

As for emotion identification in short texts, some studies have started to implement classification models in the processing of microblogs and news headlines. Strapparava and Mihalcea (2007) focused on the emotion classification of news headlines. Go, Huang, and Bhayani (2009) used distant supervision for sentiment classification of Twitter messages, or “tweets,” with SVM outperforming Naïve Bayes and Maximum Entropy and resulting in 82.2% accuracy. Sun et al. (2010) focused on the Plurk microblogging platform and used textual content and the NTU Sentiment Dictionary to build their feature set. In a recent study, Vanzo, Croce and Basili (2014) investigated contextual information found in tweets and proposed a context-based model for sentiment detection.

### 2.3.2 Adverting Legality

Computational advertising has attracted much attention in recent years. How to “best match” between a given user and a suitable advertisement in a given context is one of the major issues (Gabrilovich et al., 2008 and 2009). Some previous studies on Internet advertising focused on bidding (selecting) advertisements and placing them in the best (right) positions. Ghosh et al. (2009) proposed bidding strategies for the



allocation of advertisements. Edelman, Ostrovsky and Schwarz (2007) investigated generalized second-price (GSP) auction for online advertising. Huang, Lin, and Chen (2008) classified instant messaging dialogues into the Yahoo categories and applied the method to advertisement recommendation. Cheng and Cantú-Paz (2010) proposed a framework to predict the probability of individual users' clicks on ads. Scaiano and Inkpen (2011) used a set of Wikipedia articles as an annotated corpus to find negative key phrases to avoid displaying advertisements to non-target audience.

Unlike advertisement bidding, matching and recommendation in computational advertising, this study focuses on illegal advertisement recognition. Illegal advertising is similar to ad spams<sup>2</sup> for financial gain, but the former uses false, overstated or misleading statements to defraud customers, while the latter creates artificial ad traffic, inflates click/impression, and so on, to defraud online advertising systems such as AdWords. Advertising legality recognition is a new field in NLP and data mining. Food, cosmetics, and medicine are three major objects used in illegal advertising. Since advertisements that make health claims are highly regulated in many countries, this study focuses on food-related and cosmetics-related advertising.

### 2.3.3 Irony

Sarcasm and irony have been studied by linguistics and cognitive scientists for years (Giora and Fein, 1999; Gibbs and Colston, 2007), but there has been no concrete definition of the linguistic structure of irony. In recent years, some studies within computer science domain, as mentioned below, have also started focusing on the processing of sarcasm and irony.

---

<sup>2</sup> <http://support.google.com/adwordspolicy/bin/answer.py?hl=en&answer=50424>

Veale and Hao (2010) constructed a corpus of ironic similes using the wildcarded query “as \* as a \*” in a search engine. Lukin and Walker (2013) used a bootstrapping method to improve the performance of the classifiers for identifying sarcastic and nasty utterances in online dialogues.

In order to perform irony detection experiments on tweets, many studies used the hashtag to automatically find ironic expressions on Twitter. This is an efficient way because a Twitter user can annotate his or her ironic expressions by adding the #irony or #sarcasm hashtag to tweets. Davidov et al. (2010) collected messages from Twitter and product reviews from Amazon.com using the Mechanical Turk service. In their study, the #sarcasm hashtag was used as ground truth, and a k-nearest neighbor strategy was used for classification. González-Ibáñez et al. (2011) also made use of hashtags in Twitter as labels to build a sarcasm corpus. According to the latter paper, both human classification and automatic classification methods achieved low accuracy in sarcasm detection. Filatova (2012) used a crowdsourcing approach to generate an irony and sarcasm corpus. Reyes et al. (2012) analyzed humor and irony based on user-generated tags, such as #humor and #irony, on Twitter.

Hashtag-based approaches are not always suitable for irony corpus construction for all languages. As of March 9, 2014, only 113 messages were found to contain the hashtag #反諷 (#irony) on Weibo, the largest Chinese-language microblogging platform. This study differs from the previous work in that we employ negative emoticons and positive words as clues to capture irony. The linguistic patterns mined from the irony corpus can be used to detect if a sentence is ironic.

## Chapter 3 Emotions in Microblogs



Corpora annotated with emotion information are indispensable for emotion modeling. Social media, such as blogging platforms, encourage users to share their emotions through writing. For example, bloggers regularly use emoticons to express personal feelings in their posts. To encourage more reader interactions, some news media, e.g., Yahoo Kimo News, provide readers a voting mechanism so they can express their feelings about news articles they've just read. A collection of blogger posts and news article reader responses forms writer and reader emotion-tagged corpora, respectively, facilitating writer emotion and reader emotion mining.

Previous studies (e.g., Yang, Lin and Chen, 2007a; Yang, Lin and Chen, 2007b; Yang, Lin and Chen, 2008) have used an emotion-tagged blog corpus to investigate the ways in which people express their emotions, trying to detect writers' affective status with textual contents they have written. While these studies aimed to perform emotion analysis and detection from writers' perspective, some other papers have focused on reader emotion generation (Lin, Yang and Chen, 2007; Lin and Chen, 2008; Lin, Yang and Chen, 2008) using emotion-tagged news datasets, modeling how readers react to articles on news websites.

To study how writer emotion affects readers' feelings, Yang, Lin and Chen (2009) used the Yahoo Kimo Blog and Yahoo Kimo News to produce a dataset annotated with both writers' and readers' emotions. They constructed a document-level reader-emotion classifier using the Yahoo Kimo News corpus and applied the classifier to the Yahoo Kimo Blog corpus. In this way, a new blog corpus labeled with both writer and reader emotions was obtained.

The major problem with the above approach is that the reader emotion tagging on



the writer corpus depends on classification performance. Plurk, a unique social network and microblogging platform, provides useful resources where we can find dialogues consisting of original posts and their corresponding replies. A poster begins by publishing a post along with an emotion, then a replier responds to the post and labels it with an emotion symbol. The replier acts as a reader as well as a writer when they send their responses. Therefore, the original poster has only a writer emotion, but the replier has both a writer emotion and a reader emotion.

In this chapter, emotion mining models built from either the writer perspective, reader perspective, and combination of both writer and reader perspectives are proposed. To collect data including both writers' and readers' emotional information, messages are extracted from Plurk, ending up with 50,000 conversations in the dataset.

The Support Vector Machine (SVM) algorithm is adopted to build classifiers for emotion prediction. Like many other related studies, textual features are used as basic feature sets for training and testing. Since the conversations in the dataset contain communications and interactions between users, some non-linguistic features were also taken into account. As a result, four types of features, namely, linguistic features, social relation, user behavior, and relevance degree, are used.

### 3.1 Dataset

#### 3.1.1 Emoticons on Plurk

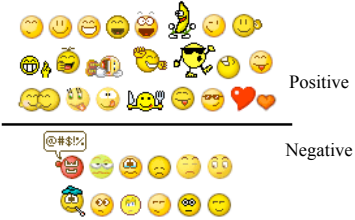


Figure 3.1 Emoticons as positive and negative labels

The short messages in the Plurk Corpus are used for the emotion identification experiments. The Plurk website provides 78 basic graphic emoticons, which are commonly used by users. In this experiment, 35 of the emoticons are chosen and categorized into the positive and negative group. The other 43 are either neutral or cannot be clearly categorized and are thus excluded to minimize uncertainty. Figure 3.1 lists the Plurk emoticons used in this study.

### 3.1.2 The Plurk Dataset

In the Plurk dataset, there are 50,000 conversations dating from Jun 21, 2008 to Nov 7, 2009, and each of them consists of a post with or without an emoticon and a corresponding reply with an emoticon. Only the earliest reply for each conversation has been included because this can help us make sure the reply is a response to the original post rather than to other responses. All messages are in Traditional Chinese.

Some messages are filtered out by their qualifiers. For example, messages with the “share” qualifier are not included because most “shares” are hyperlinks or images rather than general text messages. If a message contains an emoticon that is not shown in Figure 3.1, it will also be filtered out. Such an emoticon does not present obvious positive or negative emotion and will not be used in this study.

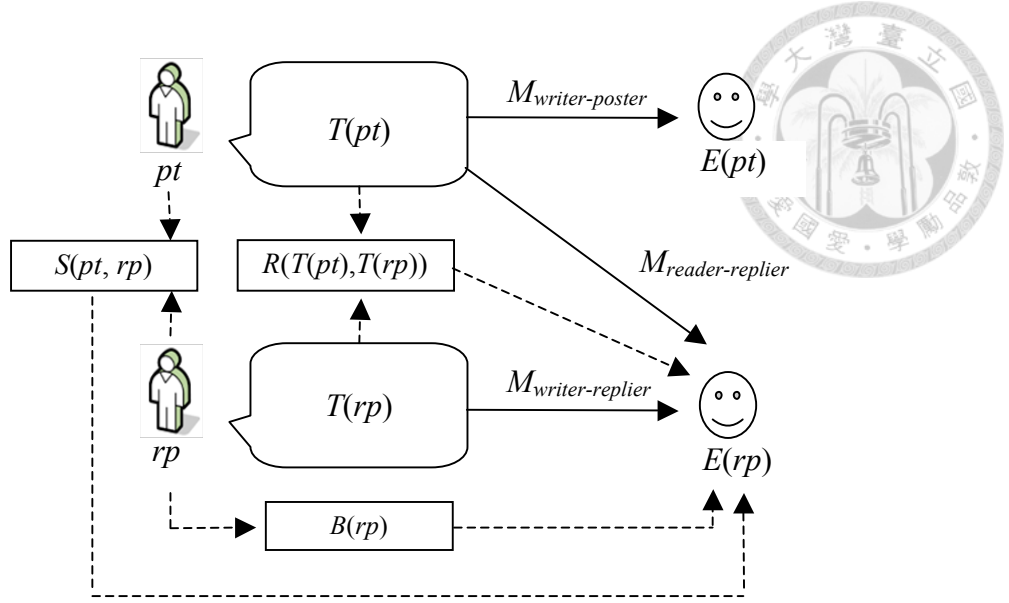


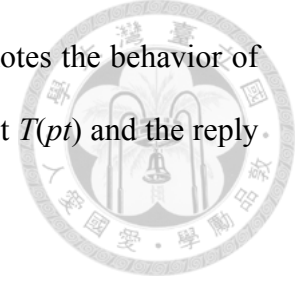
Figure 3.2 Different emotion generation models on Plurk

In the dataset, there are 42,115 conversations with a positive reply and 7,885 conversations with a negative reply. These conversations are obtained randomly from the Plurk website, and thus they should reflect their actual distribution on Plurk. For this reason the dataset is used without adjusting the proportion of the two emotion types. The proportion of positive conversations (84.23%) is used as baseline.

### 3.2 Reader/Writer Perspective

Most related studies focus on the analysis and detection of writers' emotion, since a writer's content has a more direct link to his emotion, and corpora containing writers' emotion are easier to find on the Web. In our experiment, the goal is to model the generation of reader's emotion and to see how this kind of emotion can be related to the content written by poster, replier, or both. Depending on different perspectives, we have 3 types of models: reader model, writer model, and reader + writer model. Figure 3.2 shows important components in these models: the poster  $pt$ ; the text  $T(pt)$  that  $pt$  posts; a replier  $rp$ ; the text  $T(rp)$  used by  $rp$  to reply to  $pt$ ;  $rp$ 's emotion  $E(rp)$ ;  $S(pt, rp)$ , which

denotes the social relationship between  $pt$  and  $rp$ ;  $B(rp)$ , which denotes the behavior of  $rp$ ; and  $R(T(pt), T(rp))$ , which denotes the relevance between the post  $T(pt)$  and the reply  $T(rp)$ .



### 3.2.1 Reader Perspective

By looking at a replier's emotion from readers' perspective, a reader model can be built. In this model, it is assumed that a replier's emotion is directly generated by reading the poster's message and then the replier expresses his emotion by using an emoticon in his reply. It is indicated by the model  $M_{reader-replier}$  in Figure 3.2. We therefore have  $E(rp)=M_{reader-replier}(T(pt))$  where  $M_{reader-replier}$  is a function that maps  $T(pt)$  into an emotion. Besides  $T(pt)$ , the social relationship between  $rp$  and  $pt$  and the behavior of  $rp$  such that  $E(rp)=M_{reader-replier}(T(pt), S(pt, rp), B(rp))$  are also taken into account.

### 3.2.2 Writer Perspective

In a conversation, both the poster and the replier generate textual contents. To model emotion generation from writers' perspective, it is assumed that users' emotions are related to their own contents. Thus, we have two types of writer model: poster's writer model and replier's writer model. This study focuses on replier's writer model, while poster's writer model is listed for comparison. For replier's writer model, a replier's content is used to predict his own emotion. The model  $M_{writer-replier}$  in Figure 3.2 indicates the generation of a replier's emotion from writers' perspective. This can be represented by  $E(rp)=M_{writer-replier}(T(rp))$  where  $M_{writer-replier}$  is a function that maps  $T(rp)$  into an emotion. Besides  $T(rp)$ , we can consider social relationship between  $rp$  and  $pt$  and the behavior of  $rp$  such that  $E(rp)=M_{writer-replier}(T(rp), S(pt, rp), B(rp))$ . For

poster's writer model, a post's content is used to predict his emotion. That is,  
 $E(pt)=M_{writer-poster}(T(pt))$ .



### 3.2.3 Reader and Writer Perspective

In this kind of model, readers' and writers' perspectives are combined, assuming that a replier's emotion is related to both poster's content and the replier's own content. A replier's emotion is thus predicted using poster's and replier's texts. This is represented by  $E(rp)=M_{reader-writer}(T(pt), T(rp), R(T(pt), T(rp)))$  where  $M_{reader-writer}$  is a function maps  $T(pt), T(rp), R(T(pt), T(rp))$  into an emotion. Besides textual information, social relationship between  $rp$  and  $pt$  and the behavior of  $rp$  into this function can also be incorporated.

## 3.3 Emotion Modeling

SVM is adopted as classifiers to predict emotions from reader and/or writer perspectives. Besides textual features, non-textual features, such as social relation, user behavior, and relevance degree, are also used.

### 3.3.1 Text Features (T)

Since about 70% of Chinese words are disyllabic and new words and slangs are commonly used on social media, bigrams are used instead of words as features. Chinese character bigrams in all poster's and/or replier's messages are extracted. The relationship between a bigram  $w$  and an emotion  $e$  as probability  $P(w|e)$  is modeled.

The training set is composed of conversations between posters and repliers. A conversation scenario between a poster and a replier is described as follows. A poster  $pt$  writes a post  $T(pt)$  with emotion  $E(pt)$ . After a replier  $rp$  reads the post  $T(pt)$ ,  $rp$  writes a



reply  $T(rp)$  with emotion  $E(rp)$ . Note that poster  $pt$  writes and replier  $rp$  reads the same message  $T(pt)$  and respectively express emotions  $E(pt)$  and  $E(rp)$ . In contrast, replier  $rp$  reads and writes different messages, i.e.,  $T(pt)$  and  $T(rp)$ , with the same emotion  $E(rp)$ .

As a result, three datasets can be built:  $D_{writer-poster}$ ,  $D_{reader-replier}$ , and  $D_{writer-replier}$ .  $D_{writer-poster}$  is composed of all the messages of posters along with their emotions.  $D_{reader-replier}$  consists of all the messages that repliers read and emotions that they express.  $D_{writer-replier}$  denotes a set of messages and emotions that repliers convey. The three data sets are used to train  $P_{writer-poster}$ ,  $P_{reader-replier}$ , and  $P_{writer-replier}$ , respectively.

To implement SVM in the experiments, libSVM (Chang and Lin, 2001) is adapted as the classification tool. The libSVM parameter selection tool shows that the classifiers yield the best results when  $C=3$  and  $\gamma=0.13$ .

### 3.3.2 Social Relation (S)

Each replier has his own interaction preference. The social relationship between a poster and a replier is the first cue. Therefore, the social relationship between two users is measured according to their interaction degree. The following three features are proposed.

As shown below,  $S_1$  defines the interaction degree between users  $u_1$  and  $u_2$  as their total number of interactions.  $D$  is a multiset of conversations  $(u_1, u_2)$ , and  $u_1$  and  $u_2$  are poster and replier in the conversation.

$$S_1(u_1, u_2) = \sum_{(u_1, u_2) \in D} 1 \quad (1)$$

The second social relation feature  $S_2$  considers how often user  $u_1$  posts messages.

$$S_2(u_1, u_2) = \frac{S_1(u_1, u_2)}{(end - start) \sum_{(u_1, replier) \in D} 1} \quad (2)$$

In (2), *start* and *end* denote the starting and ending day of the interaction between user  $u_1$  and  $u_2$ .  $S_2$  equals to  $S_1$  divided by the frequency of posts by poster  $u_1$ .

It is also important to take how often a replier posts a reply into consideration.  $S_3$  defined as follows captures this idea.

$$S_3(u_1, u_2) = \frac{S_1(u_1, u_2)}{(end - start) \sum_{(poster, u_2) \in D} 1} \quad (3)$$

### 3.3.3 User Behavior (B)

The user behavior features model the subjective tendency of a user. The history of a specific replier shows which emotions he tends to express.  $B_{-int}$  defines the negative tendency of user  $u$ .

$$B_{-int}(u) = \frac{C(E(u)=0)}{C(E(u)=0) + C(E(u)=1)} \quad (4)$$

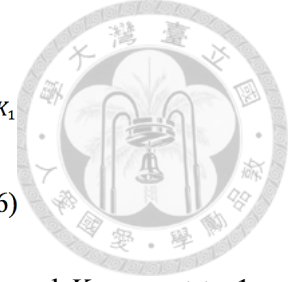
In the above formula,  $u$  is a replier,  $E(u)$  is the replier's emotion with a value 0 (negative) or 1 (positive).  $C$  is the frequency of  $E(u)$ . This indicator does not take the interaction with posters into account.

The other user behavior feature represents how often a replier expresses his positive emotion to a specific poster. This feature is called interactive behavior ( $B_{+int}$ ) and is defined as follows.

$$B_{+int}(u) = \frac{C(E(u)=1)}{\sum_{(poster, u) \in D} 1} \quad (5)$$

In some cases, replier's behavior history is not available. Back-off smoothing is used to deal with this issue. Interactive user behavior after smoothing ( $B_s$ ) is defined as:

$$B_s(E(rp) = e|T(pt)) = \begin{cases} P(E(rp) = e|pt = u) & \text{if } \sum_{e \in EM} C(E(rp) = e, pt = u) > K_1 \\ P(E(rp) = e) & \text{if } \sum_{e \in EM} C(E(rp) = e) > K_2 \\ P(E(RP) = e) & \text{otherwise} \end{cases} \quad (6)$$



In (6),  $rp$  is a replier,  $pt$  is a poster, and  $RP$  is a set of all repliers.  $K_1$  and  $K_2$  are set to 1 in the experiments.

### 3.3.4 Relevance Degree (R)

Although a post and its reply are in the same conversation, they are not necessarily on the same topic or fully related to each other. This may affect the use of emoticons, so relevance degree should be dealt with.  $R(T(pt), T(rp))$  is defined as follows:

$$R(T(pt), T(rp)) = \begin{cases} 1 & \text{if there exists an anaphoric element in } T(rp) \\ 0.5 + \frac{\text{total overlapped bigrams in } T(pt) \text{ and } T(rp)}{\text{total bigrams in } T(pt)} & \text{otherwise} \end{cases} \quad (7)$$

If there exists an anaphoric element or a conjunction in replier's message, then it is assumed that the conversation is related and the value 1 is assigned to the relevance degree. Nine anaphoric elements and 43 conjunctions are adopted. Otherwise, we check if the post and the reply overlap. More overlapped words mean that they are more related. It is assumed that the post and the reply have some basic relationship, so the default relevance degree is set to 0.5.

### 3.3.5 Normalization

The size of the linguistic feature set is much larger than the three non-linguistic feature sets, so the following vector normalization method is applied to deal with the issue:



$$F = (f_1, f_2, f_3, \dots, f_n) \quad (8)$$

$$NF = (nf_1, nf_2, nf_3, \dots, nf_n) \quad (9)$$

$$nf_i = \frac{f_i}{\sqrt{f_1^2 + f_2^2 + f_3^2 + \dots + f_n^2}} \quad (10)$$

$F$  is a vector representing the values of  $n$  features  $f_1, f_2, f_3, \dots, f_n$ . Each  $f$  value is divided by the length of  $F$ . Finally, the new values  $nf_1, nf_2, nf_3, \dots, nf_n$  constitute the normalized  $F$ .

### 3.4 Results

Classifiers were trained and tested with 10-fold cross-validation. In this section, the results of the models from the three types of perspectives are shown and discussed.

#### 3.4.1 Text Features ( $T$ )

|            |                     |               |
|------------|---------------------|---------------|
| $T$        | Reader model        | 80.67%        |
|            | Writer model        | <b>88.75%</b> |
|            | Reader+Writer model | 88.71%        |
| $S$        |                     | 82.78%        |
| $B_{-int}$ |                     | 84.14%        |
| $B_{+int}$ |                     | 86.25%        |
| $B_s$      |                     | 86.93%        |
| $R$        |                     | 81.53%        |

Table 3.1 Accuracies of different feature sets

An individual feature set is used at a time to compare the performance. The linguistic feature set ( $T$ ) is used to model replier's emotion generation from three different perspectives. When performing the prediction task with the reader model and

the writer model, 3,000 bigrams from poster's and replier's messages were used, respectively. To perform the task with the reader + writer model, all the bigrams from both the reader and writer models were used. As a result, a total of 6,000 features are used.

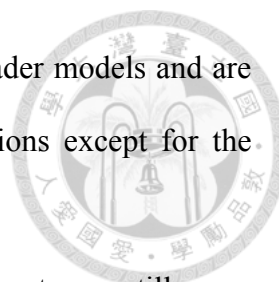
Table 3.1 shows that the writer model and the reader + writer model achieved higher performance than the reader model. The performance of the writer model is slightly higher than that of the reader + writer model, but the t-test shows that the difference is insignificant. The performance of the writer model and the reader + writer model is higher than the baseline (84.23%), while the performance of the reader model is lower than that of the baseline.

The classifier with the interactive user behavior ( $B_{+int}$ ) feature outperformed the one with the non-interactive user behavior ( $B_{-int}$ ) feature, achieving performance (86.25%) higher than the baseline. After applying back-off smoothing, the interactive user behavior ( $B_s$ ) proved to achieve even higher performance (86.93%), which is the best among all non-linguistic feature sets.

The social relation ( $S$ ) and relevance degree ( $R$ ) features result in lower performance than the baseline. In summary, when each of the non-linguistic feature sets is used individually,  $B_s$  is the most effective:  $B_s > B_{+int} > B_{-int} > S > R$ . For the behavior feature set, back-off smoothing is useful. In addition, the behavior pattern in response to a specific poster is more useful than to all posters, suggesting that the affective interaction between two given users may be based on a certain pattern.

### 3.4.2 Combination of Feature Sets

Experimentation with some combinations of different feature sets is also performed. Table 3.2 shows the results with these combinations from reader, writer, and



reader and writer perspectives. Writer models still outperformed reader models and are slightly better than reader+writer models for all feature combinations except for the model with the  $T + B_s + S$  combination.

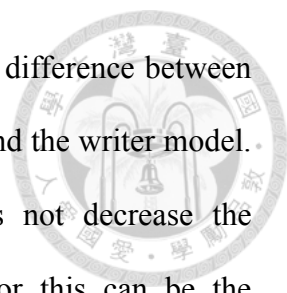
When combined with textual features, the behavioral feature set was still more powerful than social relation and relevance degree. However, all the 3 non-textual feature sets are helpful since paired t-tests show that the differences between  $T$  and  $T + B_s$ ,  $T$  and  $T + S$ , and  $T$  and  $T + R$  are significant ( $p < 0.05$ ).

|                   | Reader Models | Writer Models | Reader + Writer Models |
|-------------------|---------------|---------------|------------------------|
| $T$               | 80.67%        | 88.75%        | 88.71%                 |
| $T + S$           | 83.42%        | 89.60%        | 89.26%                 |
| $T + B_s$         | 88.02%        | 91.42%        | 91.16%                 |
| $T + R$           | 82.73%        | 89.14%        | 88.93%                 |
| $T + B_s + R$     | 88.14%        | 91.48%        | 91.27%                 |
| $T + B_s + S$     | <b>88.42%</b> | <b>91.60%</b> | <b>91.61%</b>          |
| $T + B_s + S + R$ | 88.37%        | 91.53%        | 91.30%                 |

Table 3.2 Accuracies of models with different feature combinations

Because  $B_s$  is most useful when used with textual features,  $T + B_s$ ,  $T + B_s + S$  and  $T + B_s + R$  were compared to find out how  $S$  and  $R$  can improve performance. For the reader models, the difference between  $T + B_s$  and  $T + B_s + S$  is significant ( $p < 0.05$ ), but the difference between  $T + B_s$  and  $T + B_s + R$  is insignificant. This suggests that  $T + B_s + S$  is a more useful combination than  $T + B_s + R$ . For writer and reader + writer models,  $T + B_s + S$  still outperformed  $T + B_s + R$ .

Although each of the 3 non-linguistic features can improve performance, combining all of them ( $T + B_s + S + R$ ) does not achieve the highest performance. The best performance is achieved by the combination of  $T + B_s + S$  regardless of which



perspective is adopted. According to results of the paired t-test, the difference between  $T + B_s + S + R$  and  $T + B_s + S$  is insignificant for the reader model and the writer model. This suggests that although adding  $R$  to the combination does not decrease the performance significantly, it is also not helpful. The reasons for this can be the following: both social relation and interactive behavior are related to interaction between two specific users, so their effects may overlap. In addition, only 14.73% of the conversations have a relevance value higher than 0.5.

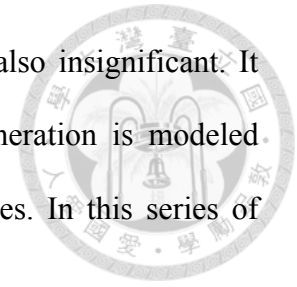
### 3.4.3 Different Perspectives

For all feature set combinations, the writer models and the reader + writer models achieve better performance than the reader models. These differences are significant according to the paired t-tests, which suggests that the message generated by the replier him- or herself contains more useful information than the message generated by the poster and then read by the replier.

When using the textual feature set only, the performance of the reader model (80.67%) was much lower than that of the writer model (88.75%) and reader + writer model (88.71%). When  $T$  is used with  $B_s$  and  $S$ , in contrast, the performance of the reader model is 88.42%, only slightly lower than the performance of the writer model (91.60%) and the reader + writer mode (91.61%). This indicates that non-linguistic features play a more important role when modeling emotion generation on a social network.

The performance of textual feature set for the writer model is 88.75%, slightly higher than that for the reader + writer model (88.71%). According to results of the paired t-test, the difference between them is insignificant. For the  $T + B_s + S$  combination, the performance of the reader + writer model (91.61%) is slightly higher

than that of the writer model (91.60%), though the difference is also insignificant. It thus makes little difference in performance whether emotion generation is modeled from writers' perspective or both readers' and writers' perspectives. In this series of experiments, 91.61% was the highest accuracy achieved.



#### 3.4.4 Writer Model

As mentioned in the Section 3.3, posters' writer mode also exists. In such a model, only the linguistic feature set can be used, and the classification accuracy is 89.19%. The t-test shows that the difference between posters' and repliers' writer models are insignificant ( $p < 0.082$ ). However, it is important to note that the dataset used for the posters' writer model differs from the one used for the repliers' writer model, so this comparison is for reference only.

### 3.5 Writer-Reader Emotion Transition

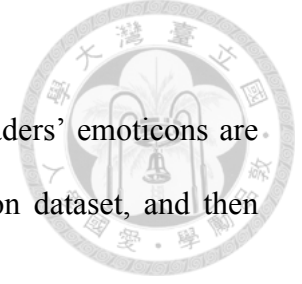
In a conversation, the emotions behind the writer's post and replier's response can be different. The process of changing from the writer emotion to reader emotion is called writer-reader emotion transition in this section. The study on what factors affect the emotion transition is important for human language understanding and has many potential applications.

Most previous studies on emotion analysis first focused on writers' perspective. Pang et al. (2002) classified movie reviews into positive and negative emotions. Wiebe (2000) investigated the subjectivity of adjectives. Aman and Szpakowicz (2007) labeled phrases with emotional categories. Beyond binary classification, Mishne (2005) classified blog posts into 37 emotion classes. Some recent studies (Lin, Yang and Chen, 2008; Lin and Chen, 2008; Yang, Lin and Chen, 2009) proposed reader emotion



analysis approaches.

In this section, Plurk conversations with both writers' and readers' emoticons are used. Linguistic features are mined from the writer-reader emotion dataset, and then emotion transitions between writers and readers are predicted.



### 3.5.1 Sentiment Word Mining

On microblogging platforms, a microblogger may play the role of both a writer and a reader. When posting the first message, he or she is a writer. After reading the repliers' messages, he or she can give some feedback. In this case, he or she becomes a reader of the replies. To avoid the confusion, we collect only the first message of a conversation and the first reply to the first message. Then the corpus is divided into four datasets based on the emotion transition types:  $(pos, pos)$ ,  $(pos, neg)$ ,  $(neg, neg)$  and  $(neg, pos)$ . For clarity, the four datasets are named as  $PP$ ,  $PN$ ,  $NN$ , and  $NP$  datasets, respectively. A total of 79,042 conversations form the experimental corpus. The number of instances in each dataset  $PP$ ,  $NN$ , and  $NP$  is 20,000. The number of instances in the dataset  $PN$  is 19,042 because fewer examples of  $(pos, neg)$  can be found.

The Yahoo word segmenter “斷章取義” was used for Chinese word segmentation and part-of-speech (POS) tagging of the datasets. Then each of the datasets was compared with the other ones to examine the differences between their word distribution.

|      | $PP$  | $PN$  | $NN$  | $NP$  |
|------|-------|-------|-------|-------|
| $PP$ | 1     | 0.899 | 0.816 | 0.871 |
| $PN$ | 0.899 | 1     | 0.922 | 0.940 |
| $NN$ | 0.816 | 0.922 | 1     | 0.953 |
| $NP$ | 0.871 | 0.940 | 0.953 | 1     |

Table 3.3 Similarity among emotion transition datasets

The function words with the POS of *article*, *preposition*, and *conjunction* were removed from the datasets. Each dataset is represented by a word vector  $(w_1, w_2, \dots, w_n)$ ,

where  $w_i$  is the normalized weight of the word  $w_i$ . The weight  $w_i$  in a dataset is the total occurrences of  $w^i$  divided by total number of words in the dataset. Cosine similarity is employed to measure the similarity of each pair of datasets. Table 3.3 shows the results.

The dataset *PP* contains no negative emotions and *NN* contains no positive emotions, so they differ a lot from each other as expected. As can be seen in Table 3.3, *PP* and *NN* have the lowest cosine similarity. The pair with the highest cosine similarity is *NN* and *NP*, showing that negative writer messages are similar regardless of reader emotions. In other words, the same writer message with negative emotion can cause either positive or negative reader emotions. This may make reader emotions more difficult to predict if the writer message is negative. In contrast, *PP* and *PN* have lower cosine similarity. It means that reader emotions are relatively easier to predict if the writer message is positive.

Generally speaking, all the datasets with any negative emotions, including *PN*, *NN* and *NP*, have higher cosine similarity ( $> 0.9$ ) when compared with each other. The similarity of *PP* and any other dataset is lower ( $< 0.9$ ).

The log relative frequency ratio, or *logRF*, of words in two datasets *A* and *B* is defined as follows. It is used to select critical features that capture the emotion transition.

For each  $w^i \in A \cup B$ , compute

$$\log RF_{AB}(w^i) = \log \frac{\frac{f_A(w^i)}{|A|}}{\frac{f_B(w^i)}{|B|}}$$

, where  $\log RF_{AB}(w^i)$  is a log ratio of relative frequencies of word  $w^i$  in *A* and *B*.  $f_A(w^i)$  and  $f_B(w^i)$  are frequencies of  $w^i$  in *A* and in *B*, respectively.  $|A|$  and  $|B|$  are total words in *A* and in *B*, respectively. The log relative frequency ratio is used to estimate the distribution of the words in datasets *A* and *B*.



The interpretations of  $\log RF_{AB}(w^i)$  are shown as follows.

- (1) If the relative frequency of  $w^i$  in  $A$  is higher than in  $B$ , then  $\log RF_{AB}(w^i) > 0$ . Those words with a positive ratio form a set  $A-B$ .
- (2) If the relative frequency of  $w^i$  in  $B$  is higher than in  $A$ , then  $\log RF_{AB}(w^i) < 0$ . Those words with a negative ratio form a set  $B-A$ .
- (3) If the relative frequency of  $w^i$  in  $A$  is approximately equal to that in  $B$ , then  $\log RF_{AB}(w^i) \cong 0$ .

$A$  and  $B$  can be one of the following combinations.

$$A=PP, B=PN$$

It captures the emotion transitions  $pos \rightarrow pos$  and  $pos \rightarrow neg$ . Those words in  $PN-PP$  may be probable to affect the emotion transitions from positive to negative. Those words in  $PP-PN$  may be probable to keep the emotion unchanged, i.e., remain in the positive state.

$$A=NP, B=NN$$

It captures emotion transitions  $neg \rightarrow pos$  and  $neg \rightarrow neg$ . Those words in  $NP-NN$  may have some effects on the emotion transition from negative to positive. Those words in  $NN-NP$  may keep the emotion unchanged, i.e., remain in the negative state.

$$A=PP \cup PN, B=NN \cup NP$$

It captures positive and negative emotion representations of *writers*. A writer emotion dictionary  $ED_W$  can be constructed based on the log relative frequency ratio of words in the positive dataset  $PP \cup PN$  and negative dataset  $NN \cup NP$ .

$$A=PP \cup NP, B=NN \cup PN$$

It captures positive and negative emotion representations of *readers*. Thus, a reader emotion dictionary  $ED_R$  can be constructed from the positive dataset  $PP \cup NP$  and negative dataset  $NN \cup PN$ .



### 3.5.2 Analysis of Mined Words

To examine the mined words, the top 200 words with higher log relative frequency ratios in *PN-PP*, *PP-PN*, *NN-NP*, and *NP-NN* were used as samples. Their semantic categories in the Chinese thesaurus *Tóngyìcílín* (同義詞詞林. Mei et al., 1982), which is abbreviated as *Cilin*, were identified. In *Cilin*, there are 12 categories labeled by letters: A. human, B. object, C. time and space, D. abstract concept, E. characteristics, F. movement, G. mental activity, H. activity, I. status, J. relation, K. particle words, and L. greeting.

The word counts in each category were calculated, and only the words that could be found in *Cilin* were analyzed. Figure 3.3 shows the distribution of the words in the *pos*→*pos* transition and the *pos*→*neg* transition. Words used in positive writer contents are more likely to get positive response, except for the categories B, F, I, J and K. The most noticeable feature is greeting words (category L) such as 掰掰 (goodbye), 早安 (good morning), 晚安 (good night), which never cause the *pos*→*neg* transition. The words causing the *pos*→*neg* transition include some words in the category K such as 難道 (dubiously), 幸好 (fortunately, which usually follows a negative expression), and 到底 (exactly). These words themselves do not contain negative emotion, but are usually used in expressions related to negative emotions.

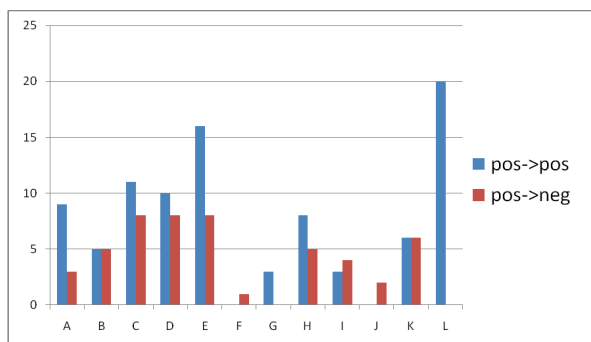


Figure 3.3 Category distribution of emotion words

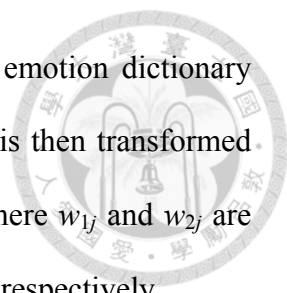
The distribution of the words used for the  $neg \rightarrow pos$  transition and  $neg \rightarrow neg$  transition is interpreted similarly. If a writer expresses negative emotion but somehow uses a greeting word in his or her message, he or she can still get a positive response from the reader in most cases. That is, the words in the group L can cause the  $neg \rightarrow pos$  transition.

The words used in the  $neg \rightarrow pos$  transition include personal status such as 好累 (tired) and 睏 (sleepy), which belongs to category I and can receive encouragement or other positive responses. As expected, the words in the transition  $neg \rightarrow neg$ , including 生氣 (angry) (category G), 恐怖 (terrible) (category G), and 可惡 (hateful) (category E), are mostly used to express negative status or characteristics.

### 3.5.3 Writer-Reader Emotion Prediction

In this experiment, the type of emotion pair (writer\_emotion, reader\_emotion) is predicted.

The training set is composed of  $m$  messages  $t_1, t_2, \dots, t_m$  sampled from Plurk. Each message  $t_i$  is annotated with an emotion pair  $(e_{i1}, e_{i2})$ , where  $e_{i1}$  and  $e_{i2}$  denote the poster's and replier's emotions, respectively. The symbol  $n$  is the number of the unique words  $w_1, w_2, \dots, w_n$  in the training set. At first, the log relative frequency ratio was



employed to build a writer emotion dictionary  $ED_W$  and a reader emotion dictionary  $ED_R$  with the approach specified in Section 3.5.1. Each message  $t_i$  is then transformed into a  $2n$ -dimension vector  $(w_{11}, w_{12}, \dots, w_{1n}, w_{21}, w_{22}, \dots, w_{2n})$ , where  $w_{1j}$  and  $w_{2j}$  are the weights of word  $w^j$  in the writer and reader emotion dictionaries, respectively.

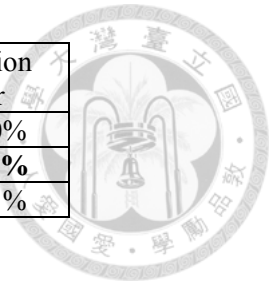
The  $m$   $2n$ -dimension vectors along with their writer-reader emotion labels were used to learn a 4-class SVM classifier. When testing, each message was represented by a  $2n$ -dimension vector, which was later used as the input to the 4-class SVM classifier. The classifier outputs the prediction, which is  $(pos, pos)$ ,  $(pos, neg)$ ,  $(neg, neg)$  or  $(neg, pos)$ .

An alternative implementation is to compose two binary SVM classifiers for writer and reader emotion prediction, respectively. Each training instance corresponds to two  $n$ -dimension vectors,  $(w_{11}, w_{12}, \dots, w_{1n})$  and  $(w_{21}, w_{22}, \dots, w_{2n})$  with writer emotion label  $e_1$  and reader emotion label  $e_2$ , respectively. The weights are determined in the same way described above. For each test message, the writer emotion prediction is made first and then the reader emotion prediction.

10-fold cross-validation is used for evaluation. Table 3.4 shows the accuracies of the 4-class SVM classifier and two binary classifiers. The baseline accuracies are 50%, 50% and 25%, since the emotion type proportions of each dataset are nearly equal. The accuracies of all proposed emotion models are higher than their corresponding baselines significantly. It shows that the sentiment word mining described in Section 3.5.1 is useful. The 4-class model outperforms the 2-class model in reader emotion prediction and writer-reader emotion pair prediction. The 2-class model outperforms the 4-class model for the writer emotion prediction task. Paired t-tests show that the performance differences between these two models for the 3 prediction tasks are all significant. The p-values are 0.017, 0.023, and 0.009, respectively.

| Prediction Task →<br>Prediction Model↓ | Writer<br>emotion | Reader<br>emotion | Emotion<br>pair |
|----------------------------------------|-------------------|-------------------|-----------------|
| Baseline                               | 50.00%            | 50.00%            | 25.00%          |
| 4-class classifier                     | 62.04%            | <b>63.85%</b>     | <b>40.86%</b>   |
| 2-class classifiers                    | <b>64.23%</b>     | 62.18%            | 38.31%          |

Table 3.4 Accuracies of emotion prediction



### 3.6 Microtexts Compared to Regular Texts

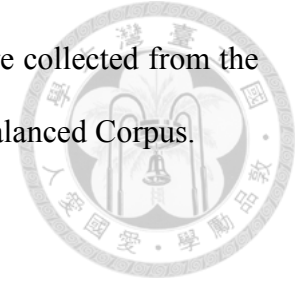
In recent years, a large number of short texts have been generated by Internet users on social media such as Twitter and Plurk. These microtexts are relatively short and informal compared with news, blogs and other online articles, making the processing of texts difficult (Ellen, 2011).

The analysis of microtext, such as short messages used in microblogs and instant messaging platforms, has started to attract researchers' attention in recent years (Ellen, 2011). To analyze the Internet informal language usages, on the other hand, Xia, Wong and Gao (2005) collected textual data from bulletin board systems and propose methods based on pattern matching and support vector machines to recognize Chinese informal words on the Internet.

To further understand the characteristics of short messages in microblogs, a comparison between a microblog corpus and a free text corpus is required. In most of the previous studies, the comparison was done by analyzing word frequency of each corpus. Kilgarriff (1998), for example, used word frequency to measure homogeneity and similarity between two corpora. Rayson and Garside (2000) adopted word frequency profiling to discover key items that differentiated one corpus from another. While this kind of comparison was based on large corpora consisting of free text corpus, Sahami and Heilman (2006) proposed a new kernel function to measure the similarity of text snippets by leveraging web search results.

This section compares microtexts with general articles by word frequency as well

as lexical semantics and sentiment analysis. The microtext data were collected from the Plurk microblogging platform and then compared with the Sinica Balanced Corpus.



### 3.6.1 Microblog Corpus vs. Balanced Corpus

The *Plurk Corpus* contains 20,265,405 Traditional Chinese posts generated by about 110,000 Plurk microbloggers from April 1 to October 31, 2009. Plurk was the most popular microblogging platform in Taiwan during that period, which means it can provide a large amount of suitable data for this study. The characteristics of the Plurk platform and the corpus are described in Section 2.2.

| Topic       | philosophy | science | society | art    | life    | literature | Total   |
|-------------|------------|---------|---------|--------|---------|------------|---------|
| #characters | 685.3K     | 102.4K  | 2761.3K | 732.2K | 1412.0K | 1278.5K    | 7892.7K |
| #words      | 451.7K     | 675.0K  | 1820.3K | 482.7K | 930.8K  | 842.8K     | 5202.8K |
| Ratio       | 8.68%      | 12.97%  | 34.99%  | 9.28%  | 17.89%  | 16.20%     | 100%    |

Table 3.5 Topic statistics of Sinica 3.0 Corpus

In comparison with the Plurk Corpus, the Sinica Balanced Corpus 3.0 (abbreviated as *Sinica Corpus*), which is a segmented and POS-tagged Chinese balanced corpus, is adopted. Sinica Corpus is composed of 5 millions words. Table 3.5 shows the topic distribution of documents in this corpus. The Plurk Corpus is segmented by maximum matching with a dictionary collected from Sinica Corpus and Sinca BOW (Bilingual Ontological Wordnet).

### 3.6.2 Analysis of Lexical Semantics

To compare these two corpora by lexical semantics, we adopt the thesaurus *Tóngyìcílín* (同義詞詞林), or *Cilin* (Mei et al., 1982). *Cilin* gathers a total of 65,464



entries of Chinese words. The word senses in *Cilin* are decomposed to a four-layer semantic structure including 12 large sense categories, 94 middle sense categories, 1,428 small sense categories, and 3,925 word sense clusters. However, it does not provide information of hypernym, hyponym, similar words, derived words, antonym, *etc.* A word with  $n$  senses corresponds to  $n$  entries so that the number of word types is less than 65,464. There are 53,644 words, including 45,586 unambiguous words and 8,058 ambiguous words. Table 3.6 shows the taxonomy of *Cilin* on the levels of large and middle sense categories. Symbols A, B, ..., L denote large sense categories and

|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>A. PERSON:</b> <b>Aa.</b> generic name, <b>Ab.</b> people of all ages and both sexes, <b>Ac.</b> posture, <b>Ad.</b> nationality/citizenship, <b>Ae.</b> occupation, <b>Af.</b> identity, <b>Ag.</b> status, <b>Ah.</b> family member, <b>Ai.</b> seniority, <b>Aj.</b> relationship, <b>Ak.</b> temperament, <b>Al.</b> ability, <b>Am.</b> religion, <b>An.</b> negative appellation                                                                                      |
| <b>B. OBJECT:</b> <b>Ba.</b> generic name, <b>Bb.</b> shape, <b>Bc.</b> part of object, <b>Bd.</b> celestial body, <b>Be.</b> terrain, <b>Bf.</b> meteorological phenomena, <b>Bg.</b> natural substance, <b>Bh.</b> plant, <b>Bi.</b> animal, <b>Bj.</b> microorganism, <b>Bk.</b> body, <b>Bl.</b> secretions/excretions, <b>Bm.</b> material, <b>Bn.</b> building, <b>Bo.</b> machine and tool, <b>Bp.</b> appliance, <b>Bq.</b> clothing, <b>Br.</b> edible/medicine/drug  |
| <b>C. TIME AND SPACE:</b> <b>Ca.</b> time, <b>Cb.</b> space                                                                                                                                                                                                                                                                                                                                                                                                                    |
| <b>D. ABSTRACT THING:</b> <b>Da.</b> event/circumstance, <b>Db.</b> reason/logic, <b>Dc.</b> look, <b>Dd.</b> function/property, <b>De.</b> character/talent, <b>Df.</b> consciousness, <b>Dg.</b> analogy, <b>Dh.</b> imaginary thing, <b>Di.</b> society/politics, <b>Dj.</b> economy, <b>Dk.</b> culture and education, <b> Dl.</b> disease, <b>Dm.</b> organization, <b>Dn.</b> quantity/unit                                                                              |
| <b>E. CHARACTERISTICS:</b> <b>Ea.</b> appearance, <b>Eb.</b> phenomenon, <b>Ec.</b> color/taste, <b>Ed.</b> property, <b>Ee.</b> virtue, <b>Ef.</b> circumstance                                                                                                                                                                                                                                                                                                               |
| <b>F. MOTION:</b> <b>Fa.</b> motion of hands, <b>Fb.</b> motion of legs, <b>Fc.</b> motion of head, <b>Fd.</b> motion of the whole body                                                                                                                                                                                                                                                                                                                                        |
| <b>G. MENTAL ACTIVITY:</b> <b>Ga.</b> mental status, <b>Gb.</b> mental activity, <b>Gc.</b> capability and willingness                                                                                                                                                                                                                                                                                                                                                         |
| <b>H. ACTIVITY:</b> <b>Ha.</b> political activity, <b>Hb.</b> military activity, <b>Hc.</b> administrative management <b>Hd.</b> production, <b>He.</b> economical activity, <b>Hf.</b> communications and transportation, <b>Hg.</b> Education/hygiene/research, <b>Hh.</b> recreation and sport, <b>Hi.</b> social activity, <b>Hj.</b> life, <b>Hk.</b> religious activity, <b>Hi.</b> superstitious activity, <b>Hm.</b> police and judicature, <b>hn.</b> wicked behavior |
| <b>I. PHENOMENON AND CONDITION:</b> <b>Ia.</b> natural phenomenon, <b>Ib.</b> physiological phenomenon, <b>Ic.</b> facial expression, <b>Id.</b> object status, <b>Ie.</b> situation, <b>If.</b> circumstance, <b>Ig.</b> beginning and end, <b>Ih.</b> change                                                                                                                                                                                                                 |
| <b>J. RELATION</b> <b>Ja.</b> association, <b>Jb.</b> similarity and dissimilarity, <b>Jc.</b> coordination, <b>Jd.</b> existence, <b>Je.</b> influence                                                                                                                                                                                                                                                                                                                        |
| <b>K. AUXILIARY:</b> <b>Ka.</b> adverb, <b>Kb.</b> preposition <b>Kc.</b> conjunction, <b>Kd.</b> particle, <b>Ke.</b> interjection, <b>Kf.</b> Onomatopoeia                                                                                                                                                                                                                                                                                                                   |
| <b>L. GREETING</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                             |

Table 3.6 Taxonomy of *Cilin*

symbols Aa, Ab, Ac, ..., Ba, ... denote middle sense categories.

Chen, Lin and Lin (2002) sampled documents from different categories of the Sinica Corpus including philosophy (10%), science (10%), society (35%), art (5%), life (20%) and literature (20%). There were 35,921 words in the test corpus. They reported the accuracies 52.85% and 34.35% for tagging ambiguous words and unknown words,

respectively, when 1,428 sense categories of *Cilin* is adopted. If unambiguous instances are also counted, the sense tagger achieved 76.04% performance.

In this study, word senses were not disambiguated. In stead, it was postulated that an ambiguous word consisting of  $n$  senses contributes each of its sense categories equally, i.e.,  $1/n$  times. Then these steps were followed: (1) count the frequency of each Traditional Chinese word in the Plurk and Sinica corpora, (2) select the top 100 and top 3,000 high-frequency words, (3) determine the sense categories of the selected words, (4) count the occurrences of each sense category, and (5) compare the distribution of their senses in each corpus. For comparability, the occurrences of each sense category in each corpus are normalized by total occurrences of all the sense categories in the same corpus. As a result, the normalized value of each sense category ranges from 0 to 1.

The union of the top 100 high-frequency words in the Plurk Corpus and those in the Sinica corpus formed a set of 140 words.  $\log RF$ , which is described in 3.5.1, is used to distinguish which words are more critical in a corpus. Using the following equation

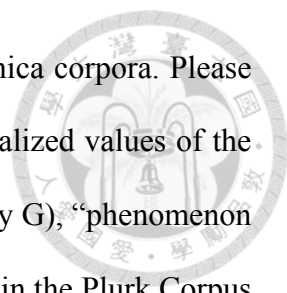
$$\log RF_{AB}(w^i) = \log \frac{\frac{f_A(w^i)}{|A|}}{\frac{f_B(w^i)}{|B|}}$$

, the following two  $\log RF$  values are calculated:

- (1)  $\log RF_{Plurk-Sinica}(w^i)$ , where  $A=Plurk$ ,  $B=Sinica$
- (2)  $\log RF_{Sinica-Plurk}(w^i)$ , where  $A=Sinica$ ,  $B=Plurk$

If the likelihood ratio  $\log RF_{Plurk-Sinica}(w^i) > 0$ , then  $w^i$  tends to occur more frequently in the Plurk Corpus than in the Sinica Corpus. Similarly, if the likelihood ratio  $\log RF_{Sinica-Plurk}(w^i) > 0$ , then  $w^i$  tends to occur more frequently in the Sinica Corpus than in the Plurk Corpus. If  $\log RF(w^i) \cong 0$ , then the relative frequencies of  $w^i$  in both corpora are approximately equal.

Figure 3.4 (on page 38) shows the sense distribution in terms of 12 large sense



categories of the top 100 high-frequency words in the Plurk and Sinica corpora. Please refer to Table 3.6 for the meanings of the sense symbols. The normalized values of the “motion” sense (category F), “psychological activity” sense (category G), “phenomenon and condition” sense (category I), and “greeting” sense (category L) in the Plurk Corpus are two times larger than those of the corresponding senses in Sinica Corpus. In contrast, the normalized values of the “person” sense (category A) and “abstract thing” sense (category D) in Sinica Corpus are two times larger than those of the corresponding senses in the Plurk Corpus.

In terms of 100 most frequent words, the Plurk Corpus has more words for “activity” (category H), while the Sinica Corpus has more words for “abstract thing” (category D). Most of the H-category words in the Plurk Corpus are related to personal experience, such as “eat” (吃), “buy” (買), and “sleep” (睡). Both the Plurk and Sinica corpora contain many auxiliaries (category K). The relative frequency of this category does not differ a lot in the two corpora. When we look into the subcategories, however, much more words in category Kd appear in the Plurk Corpus. That is, Plurk uses more words like *a* (啊), *ma* (嗎), and *la* (啦), which are sentence-final particles in Chinese. Table 3.7 lists the top 10 high  $\log RF_{Plurk-Sinica}$  words and the top 10 high  $\log RF_{Sinica-Plurk}$  words for reference.

| Ranking | logRF <sub>Plurk-Sinica</sub> | English                                     | Category | logRF <sub>Sinica-Plurk</sub> | English                              | Category   |
|---------|-------------------------------|---------------------------------------------|----------|-------------------------------|--------------------------------------|------------|
| 1st     | 一個                            | <i>yíge</i><br>(indefinite article)         | Dn       | 其                             | <i>qí</i><br>(3rd-person possessive) | Aa, Ba, Ed |
| 2nd     | 早安                            | good morning                                | L        | 表示                            | express;<br>indicate                 | Hi, Ja     |
| 3rd     | 嘍                             | <i>pū</i><br>(referring to a Plurk message) | Kf       | 並                             | and;<br>also                         | Ka, Kc     |
| 4th     | 阿                             | <i>a</i><br>(sentence-final particle)       | Hi       | 則                             | <i>zé</i><br>(conjunction)           | Ja         |
| 5th     | 晚安                            | good evening or<br>good night               | L        | 各                             | each                                 | Dn         |
| 6th     | 超                             | super                                       | Ka       | 由                             | from                                 | Kb         |
| 7th     | 囉                             | <i>Lou</i><br>(sentence-final particle)     | Kd       | 所                             | <i>suǒ</i><br>(particle)             | Dn, Kd     |
| 8th     | 耶                             | <i>Ye</i><br>(sentence-final particle)      | Kd       | 每                             | every                                | Ed         |
| 9th     | 睡                             | sleep                                       | Fd, Hj   | 及                             | and                                  | Kc         |
| 10th    | 睡覺                            | sleep                                       | Hj       | 於                             | <i>yú</i><br>(preposition)           | Kb         |

Table 3.7 List of top 10 high logRF<sub>Plurk-Sinica</sub> words and top 10 high logRF<sub>Sinica-Plurk</sub> words

The sense distributions of the top 3,000 high-frequency words in both corpora are also compared. Figure 3.5 shows that the tendency is quite similar. In the Plurk Corpus, the normalized values of the “motion” sense (category F), “phenomenon and condition” sense (category I), and “greeting” sense (category L) are still two times larger. The normalized values of the “object” sense (category B) in the Plurk Corpus become much larger than those in Sinica Corpus, while the difference of “psychological activity” sense (category G) between these two corpora becomes smaller in the top 3,000 high-frequency words. Comparatively, the value of the “abstract things” sense (category D) in the Sinica Corpus is two times larger than that in the Plurk Corpus.

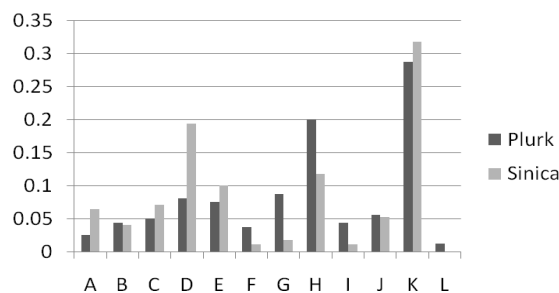


Figure 3.4 Sense distribution in terms of 12 large categories of the top 100 words

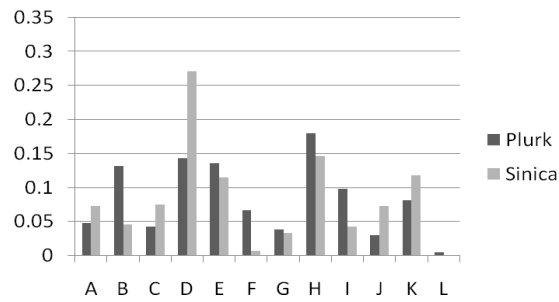


Figure 3.5 Sense distribution in terms of 12 large categories of the top 3,000 words

We can also extend this discussion from the coarsest sense set (i.e., 12 large sense categories) to the finer sense set (i.e., 94 middle sense categories). Figure 3.6 shows the sense distribution of the top 3,000 high-frequency words in Plurk and Sinica corpora. The normalized values of the senses “shape” (subcategory Bb), “motion of hands” (subcategory Fa), “motion of legs” (subcategory Fb), “motion of head” (subcategory Fc), “motion of the whole body” (subcategory Fd), “natural phenomenon” (subcategory Ia), “physiological phenomenon” (subcategory Ib), “facial expression” (subcategory Ic), “object status” (subcategory Id), “circumstance” (subcategory If), and “greeting” (subcategory L) are two times larger in the Plurk Corpus. In contrast with the Plurk Corpus, the Sinica Corpus owns two times more senses including “reason/logic” (subcategory Db), “look” (subcategory Dc), “function/property” (subcategory Dd), “society/politics” (subcategory Di), and “organization” (subcategory Dm).

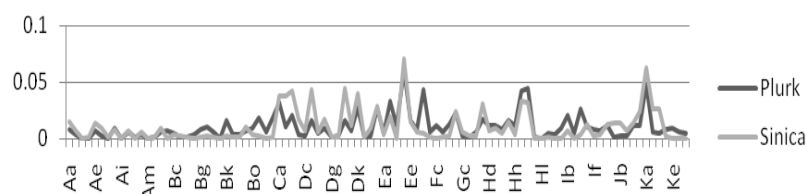


Figure 3.6 Sense distribution in terms of 94 middle categories of the top 3,000 words

### 3.6.3 Literal Text vs. Spoken Language

In this study, the average numbers of syllables of the top 100 high-frequency words in both corpora were also compared. The average word length in syllable number is 1.29 for the Plurk corpus, while that for the Sinica Corpus is 1.14. The difference between them is significant ( $p < 0.01$ ). About 86% of the top 100 high-frequency words in the Sinica Corpus are monosyllabic, while only 71% of the top 100 high-frequency words in the Plurk corpus are monosyllabic.

The above observations and Chao's (1968) statement that Classical Chinese (文言) has more mono-syllabic words and less compounds suggest that the difference between the two corpora is similar to that between Modern Chinese (白話) and Classical Chinese. Unlike general spoken Chinese, however, the Plurk Corpus also contains some Internet-specific terms, such as *pū* (噗, referring to a Plurk message), *ān'ān* (安安, a greeting word commonly used by Internet users), *shàng'àn* (上岸, referring to the action of getting offline).

### 3.6.4 Analysis of Sentiment Polarity

#### **NTU Sentiment Dictionary**

To compare the sentiment distribution in the Plurk and Sinica corpora, the National Taiwan University Sentiment Dictionary (Ku, 2007), or NTUSD, is used to classify the words in the two corpora. Table 3.8 shows the distribution of the positive, negative and neutral words in NTUSD. The most frequent type is negative words (49.786%), while the least frequent is neutral words (4.136%).

|            | Positive | Negative | Neutral |
|------------|----------|----------|---------|
| #Words     | 21,055   | 22,750   | 1,890   |
| Proportion | 46.077%  | 49.786%  | 4.136%  |



Table 3.8 Distribution of sentiment words in NTUSD sentiment dictionary

### Sentiment Analysis

Sentiment word frequency is calculated for each corpus based on the occurrence of each word in NTUSD. Figure 3.7 shows the ratio of the sentiment word frequency of the Plurk corpus to that of Sinica Corpus. For the top 100 high-frequency words, the sentiment word frequency in the Plurk corpus is 4 times larger than that in Sinica Corpus. This suggests that the most frequently used words on the Plurk Corpus are related to informal social conversations with emotion expressions. On the other hand, the top 100 high-frequency words of Sinica Corpus contain many function words. Figure 3.7 also shows that the more words are collected, the lower the ratio is. One of the reasons for this effect may be that the Plurk Corpus contains many informal and topical terms that cannot be found in NTUSD.

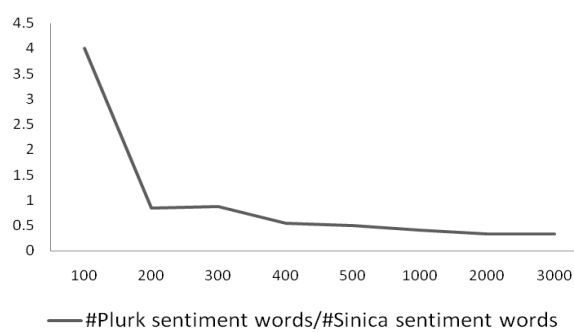


Figure 3.7 Ratio of sentiment words in top  $n$  high-frequency words in the Plurk and Sinica corpora

Figure 3.8 shows the proportions of positive, negative and neutral words in the Plurk Corpus. In the top 100 high-frequency words, much more positive words are used than negative words. As more words are taken into account, the ratio of negative words

increases and that of neutral words decreases. It illustrates that people commonly shows nonnegative feelings in a microblogging platform with social network characteristics.

As shown in Figure 3.8, the Sinica Corpus contains more positive words and less negative words. Although the ratio of neutral words is lower than that of positive words, i.e., only 4.136% of sentiment words in the NTUSD dictionary are neutral words, the ratio of neutral words is larger than that of negative words. This suggests that neutral words play an important role in the Sinica Corpus.

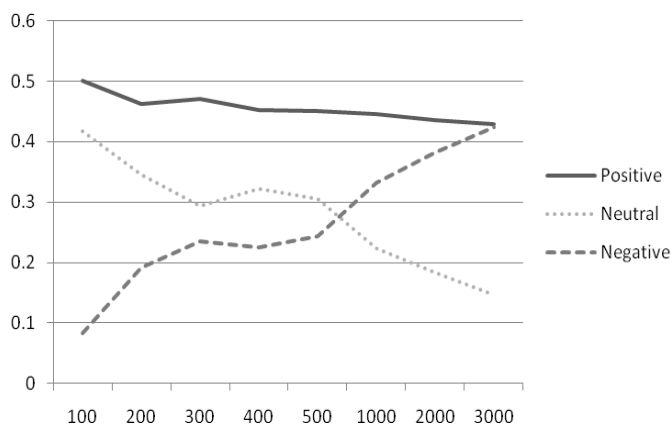


Figure 3.8 Ratio of sentiment words in top  $n$  high-frequency words in the Plurk corpus

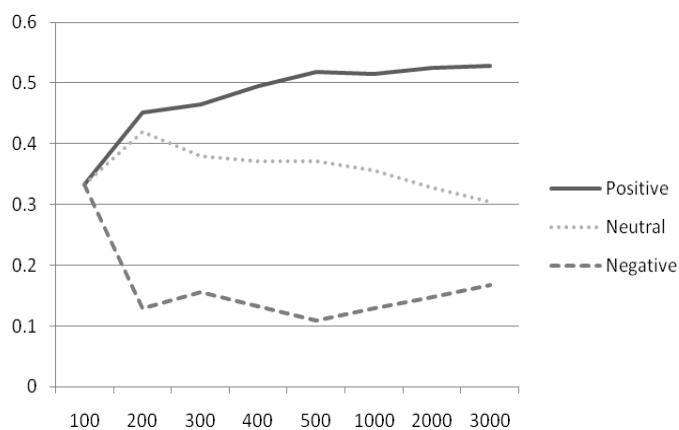


Figure 3.9 Ratio of sentiment words in top  $n$  high-frequency words in Sinica Corpus



## Chapter 4 Advertising Legality



Internet advertising has become an important marketing strategy. As online marketing continues to grow on the Internet, how consumers, advertisers and the authorities identify false and overstated advertisements becomes a major issue. In this chapter, we propose classification models to detect illegal advertisements using microtexts consisting of illegal statements made public by the government. The experimental results show that the log relative frequency ratio can be used as weights for unigram features to achieve the best accuracy. Log relative frequency ratio was also used to mine verb phrases present in illegal advertising. The classification models and the mined verb phrases were then used to construct a false online advertisement recognition system.

### 4.1 Illegal Advertising

Two forms of advertising regulation, namely statutory regulation and self-regulation, exist to protect consumers from fraudulent and misleading advertising (FTC, 2000; CFIA, 2010; DOH, 2009). Under the food and cosmetic advertising regulations of Taiwan, food-related and cosmetic-related advertisements cannot be false, overstated or misleading and should not mention any curative effects. Advertising statements that violate the regulations are called *illegal statements* in this chapter. Advertisements containing illegal statements are regarded as illegal advertisements. Examples of illegal statements provided by the government are shown in Figure 4.1.

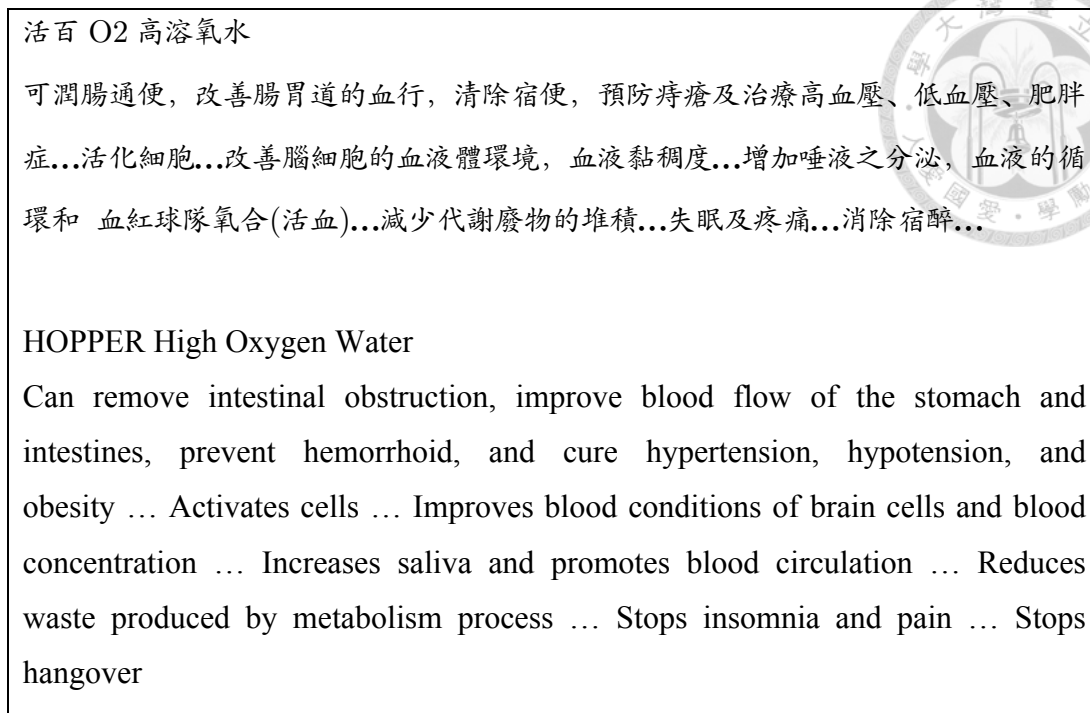


Figure 4.1 Illegal food advertising statements from an advertisement provided by the government and their English translations

Not only consumers need to identify and avoid illegal advertisements. Advertisers, governments and websites that offer advertising platforms can also benefit from automatic illegal advertisement recognition (*IDR*). Advertisers need to avoid legal issues while maximizing the effectiveness of their advertising. The authorities have to examine advertisements to decide which are allowed to be presented to Internet users. This requires a lot of time and effort. An advertising legality recognition system not only saves human effort, but also reports the illegal advertisements in real time. Websites that deliver marketing messages from other companies may also want to show only truthful advertisements to their users and avoid illegal and misleading contents. Even bloggers and auction website sellers may need to deal with legal issues. Texts and images from their websites and auctions can mention specific products and will thus be regarded as online advertising by the authority.



## 4.2 Datasets

The first step for advertising legality recognition is to obtain advertisements with appropriate legality labels. Since advertising legality can only be determined by the authority, official announcements regarding illegal advertisements must be obtained. In this study, the illegal food and cosmetic advertisement lists made public by the Taipei City Government<sup>3</sup> from July 2009 to November 2011 were collected. Each item in the list contains a product name and the corresponding illegal advertising statements. Legal statements have been removed and denoted by “...” in an item. All these statements are in Traditional Chinese.

The examples in Figure 4.1 display the fact that the government prohibits the use of statements related to curative effects and improvement of physical conditions. They also show that most illegal statements listed by the government are verb phrases consisting of a verb and an object noun. According to the observed patterns, we propose methods to expand these terms and find similar phrases in the datasets, as described in Section 4.4. This can improve the recognition tasks and help the authorities and advertisers to find problematic expressions.

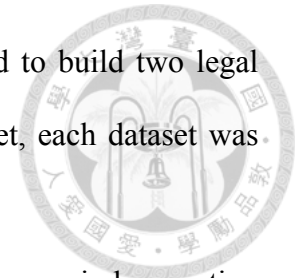
Since the government website does not provide samples of legal advertisements, we need to collect legal advertising data from other sources in order to train binary classifiers. An online shopping website in Taiwan<sup>4</sup> was used to collect legal food and cosmetic advertising documents. These data were used on the assumption that most of these advertisements comply with the advertising regulations. The data were then examined by humans to make sure that unsuitable data are removed. Product

---

<sup>3</sup><http://www.health.gov.tw/Portals/0/%E8%97%A5%E7%89%A9%E9%A3%9F%E5%93%81%E8%99%95/10010food.pdf>

<sup>4</sup> <http://www.7net.com.tw>

descriptions from the food and cosmetic advertisements were used to build two legal advertising datasets, FOOD and COS. To obtain a balanced dataset, each dataset was collected from all related product categories listed on the website.



All data were separated into sentences by punctuations, including periods, question marks, and exclamations. Only sentences with more than 3 characters were collected. Any expressions containing only product names were filtered out because product names could not be used to determine its legality. All sentences are in Traditional Chinese. To perform Chinese word segmentation and part-of-speech tagging, the CKIP segmentation and POS tagging system was adopted.<sup>5</sup>

As a result, we have four datasets for legal food advertising, illegal food advertising, legal cosmetic advertising, and illegal cosmetic advertising. For clarity, they are named as FOOD\_LEGAL, FOOD\_ILLEGAL, COS\_LEGAL, and COS\_ILLEGAL. The numbers of instances in the four datasets are 5,059, 7,033, 10,520, and 11,381, respectively.

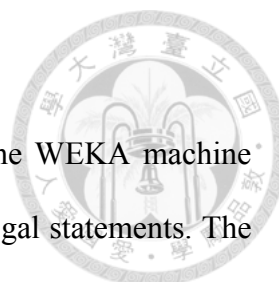
### 4.3 Recognition Models

Since the government provides only illegal advertising examples, and no official legal advertising data can be obtained, one-class classifiers are required to train the models with only the ILLEGAL datasets. In the evaluation stage, the LEGAL datasets are still required, so they are used along with the ILLEGAL datasets for testing.

Two-class classifiers were also trained using both ILLEGAL and LEGAL datasets and then compared with the one-class classifiers.

---

<sup>5</sup> <http://ckipsvr.iis.sinica.edu.tw/>



### 4.3.1 One-Class Classifiers

The OneClassClassifier module (Hempstalk et al., 2008) in the WEKA machine learning tool was adopted to train one-class classifiers with only illegal statements. The OneClassClassifier module provides a rejection rate parameter for adjusting the threshold between target and non-target instances. The target class, which corresponds to the illegal class in this study, is the single class used to train the classifier. Higher rejection rate means that more legal statements will be preferred, but more illegal statements may be incorrectly classified as legal. Naïve Bayes and Bagging classifiers were used as the classification algorithms because they achieve best performance among the algorithms that have been explored in this experiment.

Each instance in the dataset, i.e., a sentence, is represented by a word vector  $(w_1, w_2, \dots, w_{1000})$ , where  $w_i$  is a binary value indicating whether a word occurs in the sentence or not. The word list was selected from the illegal advertising datasets. To properly filter out common words, the top 1,000 high-frequency words in the Sinica Balanced Corpus of Modern Chinese<sup>6</sup> were removed from the vocabulary. The remaining top 1,000 words were then used in the vector representation.

| Rejection rate |         | 0.4    | 0.5    | 0.6    | 0.7    | 0.8    | 0.9    |
|----------------|---------|--------|--------|--------|--------|--------|--------|
| Naïve Bayes    | Illegal | 85.33% | 82.39% | 79.01% | 74.49% | 68.17% | 59.14% |
|                | Legal   | 31.07% | 39.81% | 53.40% | 63.11% | 72.82% | 86.41% |
| Bagging        | Illegal | 92.78% | 88.49% | 84.65% | 74.94% | 69.07% | 0.23%  |
|                | Legal   | 3.88%  | 17.48% | 27.18% | 65.72% | 82.52% | 99.77% |

Table 4.1 Accuracies of classifiers in different rejection rates

<sup>6</sup> <http://app.sinica.edu.tw/kiwi/mkiwi/>

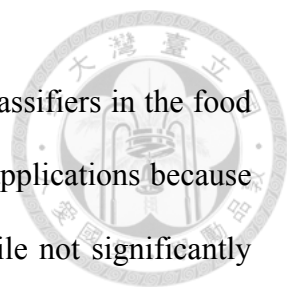
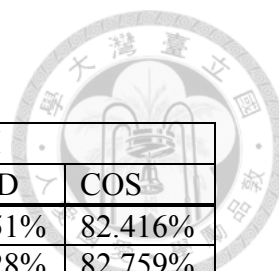


Table 4.1 shows the accuracies of Naïve Bayes and Bagging classifiers in the food dataset. The rejection rates from 0.7 to 0.8 are preferable for most applications because they result in higher accuracy for legal statement classification while not significantly reducing the performance of illegal statement detection. The 0.7 rejection rate produces high performance for the illegal class while 0.8 rejection rate does better for the legal class. The actual choice of rejection rate depends on the demands of users. For an advertiser, it is important to avoid all possible illegal statements, and a lower rejection rate will thus be more suitable. If the system is used by the authorities, a rejection rate higher than 0.7 may be preferable because they don't misjudge too many legal advertisements.

#### 4.3.2 Binary Classifiers

A one-class classifier can be used to address the issue caused by the absence of legal advertising data, but it also makes the decision boundary unfixed and unclear. To resolve this problem, both illegal and legal datasets are used to construct two-class (or binary) classifiers.

In the development processes, Naïve Bayes classifiers and SVM classifiers were implemented with libSVM (Chang & Lin, 2001). All training and test processes were based on 10-fold cross validation, and every training model was tuned with the optimized parameters to achieve the best performance. Accuracy was adopted as the evaluation metric. Table 4.2 shows the experimental results. Two kinds of classification models (Naïve Bayes and SVM) with different feature settings were explored in regard to the food (FOOD) and cosmetic (COS) datasets.



| Classification Models →<br>Features ↓      Materials → | Naïve Bayes    |                | SVM            |                |
|--------------------------------------------------------|----------------|----------------|----------------|----------------|
|                                                        | FOOD           | COS            | FOOD           | COS            |
| Unigram                                                | 89.148%        | 81.357%        | 88.851%        | 82.416%        |
| Unigram + CILIN                                        | 88.950%        | 81.311%        | 89.728%        | 82.759%        |
| Unigram + DOH                                          | 89.182%        | 81.553%        | 89.554%        | 83.658%        |
| Unigram + CILIN + DOH                                  | 89.000%        | 81.439%        | 89.727%        | 83.325%        |
| Unigram + logRF                                        | <b>90.695%</b> | <b>85.179%</b> | <b>93.433%</b> | <b>86.037%</b> |

Table 4.2 Accuracies of advertising legality recognition models

| Type | Sub-type                                                                                                                         | Example                                                                              |
|------|----------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
| 1    | 宣稱預防、改善、減輕、診斷或治療疾病或特定生理情形<br>(Claim of prevention, improvement, reduction, diagnosis or cure of diseases or physical conditions) | 減輕過敏性皮膚病<br>(reduce allergic skin disease)                                           |
|      | 宣稱減輕或降低導致疾病有關之體內成分<br>(Claim of elimination of substances that cause diseases)                                                   | 解肝毒, 降肝脂<br>(remove poison and fat in liver)                                         |
|      | 宣稱產品對疾病及疾病症候群或症狀有效<br>(Claim of effectiveness to diseases and symptoms)                                                          | 消除心律不整<br>(stop arrhythmia)                                                          |
|      | 涉及中藥材之效能者<br>(Related to effects of Chinese medicine)                                                                            | 補腎<br>(improve health condition of kidney)                                           |
|      | 引用或摘錄出版品、典籍或以他人名義並述及醫藥效能<br>(Reference to publications, books or statements by others with medical effects)                      | 「本草綱目」記載：黑豆可止痛<br>(according to the book “Bencao Gangmu,” black beans can ease pain) |
| 2    | 涉及生理功能者<br>(Related to physiological functions)                                                                                  | 分解有害物質<br>(decompose toxicants)                                                      |
|      | 涉及五官臟器者<br>(Related to organs)                                                                                                   | 增加血管彈性<br>(increase elasticity of blood vessel)                                      |
|      | 涉及改變身體外觀者<br>(Related to change of appearance of human body)                                                                     | 防止老化<br>(prevent aging)                                                              |
|      | 引用本署衛署食字號或相當意義詞句者<br>(Reference to DOH permission numbers or related expressions)                                                | 通過衛生署配方審查<br>(formula has been approved by DOH)                                      |

Table 4.3 Illegal advertising statement examples announced by the government



### **Feature Set 1: Unigrams**

In this experiment, unigrams were used as a fundamental feature set. The top 1,000 high-frequency words were selected from the legal and illegal advertising datasets as features. Only content words including verbs, nouns and adjectives were included in order to remove the words that may not be relevant. Every sentence separated by punctuations forms an instance of the datasets, and each instance is represented by a word vector  $(w_1, w_2, \dots, w_{1000})$ , where  $w_i$  is a binary value indicating whether a word occurs in the sentence or not. As seen in the “Unigram” row of Table 4.2, the accuracies of Naïve Bayes classifiers and SVM classifiers on FOOD and COS datasets are (89.148%, 81.357%) and (88.851%, 82.416%), respectively. Bigram features were also tested, but the performance is lower than unigrams. Thus, the results of bigram features are omitted here.

### **Feature Set 2: Health-Related Terms**

Advertising regulations have been announced by the government along with illegal advertising statement examples for advertisers’ reference. Table 4.3 shows some illegal examples for food-related regulations. The first type listed in the first column denotes mention of any curative effects and the second type denotes false, overstated or misleading cases. Several subtypes along with the corresponding examples are listed in the second and third columns, respectively.

According to advertising regulations, advertisers are not allowed to mention any curative effects on food and cosmetic advertisements. In order to increase the coverage of the feature sets, the words related to curative effects were expended by using a thesaurus. These statements were used as auxiliary features and combined with unigram features. The two kinds of auxiliary features shown below are used:



(1) All verbs related to curative effects in the Chinese thesaurus *Tongyicicilin* (Mei et al., 1984): This feature set is called CILIN in Table 4.2.

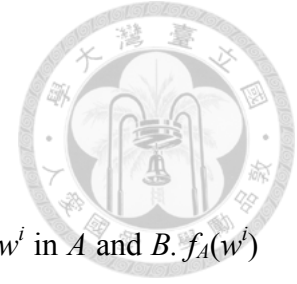
(2) Illegal statement examples listed by Department of Health (DOH) of Taiwan: This feature set is called DOH in Table 4.2.

The forth, fifth and sixth rows of Table 4.2 show the accuracies of the above feature sets. Thesaurus expansion (Unigram + CILIN) has positive effects in the SVM classifiers. Compared with the pure unigram feature sets, integrating features selected from illegal advertising statement examples of DOH (Unigram + DOH) is also useful. However, the accuracy is not further improved when all the three kinds of features are combined (Unigram + CILIN + DOH). A possible reason is that the number of terms in the CILIN feature set is high. In addition, a thesaurus always tries to collect as many terms as possible, and many uncommon words have been thus unnecessarily included in this experiment. The DOH feature set includes lists that are edited by professionals in the government, so it captures illegal advertising statements that are in actual use. However, the coverage is an issue. Section 4.4 discusses how to expand this list.

### **Feature Set 3: Log Relative Frequency Ratio**

The relative frequency ratio between two datasets has been shown to be useful to discovering collocations that are characteristic of a dataset when compared with the other dataset (Damerau, 1993). As described in Chapter 3, it can be used to model emotion transition between writers and readers. In this experiment, this idea was extended to select critical features that capture the legality transition. The log relative frequency ratio, or *logRF*, of words in two datasets *A* and *B* are defined as follows. For each  $w^i \in A \cup B$ , compute

$$\log RF_{AB}(w^i) = \log \frac{\frac{f_A(w^i)}{|A|}}{\frac{f_B(w^i)}{|B|}}$$

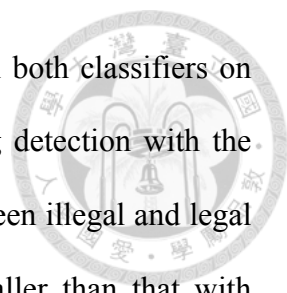


, where  $\log RF_{AB}(w^i)$  is the log ratio of relative frequencies of word  $w^i$  in  $A$  and  $B$ .  $f_A(w^i)$  and  $f_B(w^i)$  are frequencies of  $w^i$  in  $A$  and in  $B$ , respectively.  $|A|$  and  $|B|$  are total words in  $A$  and in  $B$ , respectively. The log relative frequency ratios are used to estimate the distribution of the words in datasets  $A$  and  $B$ .

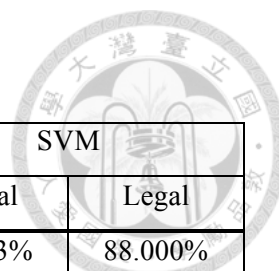
In the experiments for food advertising,  $A=FOOD\_LEGAL$  and  $B=FOOD\_ILLEGAL$ . As for the experiments for cosmetic advertising,  $A=COS\_LEGAL$  and  $B=COS\_ILLEGAL$ . The log relative frequency ratio was used as weight of each unigram in a dataset. Each sentence in the datasets is represented by a vector  $(w_1, w_2, \dots, w_n)$ , where  $w_i$  is the weight of  $i^{\text{th}}$  word from the unigram feature set. The seventh row of Table 4.2 lists the accuracy of the log relative frequency ratio feature set for the FOOD and COS advertising legality classification. The performance of both Naïve Bayes and SVM classifiers with Unigram + logRF feature settings are higher than those with the unigram and the auxiliary feature settings regardless of the type of datasets. The differences of accuracies between Unigram + logRF and all the other feature settings for both datasets are statistically significant ( $p < 0.01$ ).

## Discussion

Table 4.4 and Table 4.5 show the experimental results of Naïve Bayes and SVM classifiers with different feature settings on food and cosmetic datasets, respectively. We have the following observations. Firstly, Unigram+CILIN does not improve the accuracy of Unigram. The *Cilin* thesaurus contains many words that are not commonly used. In addition, its purpose is to help people find similar and related words conveniently. Its organization of lexical terms may not be suitable for the classification



tasks. Secondly, the accuracies of illegal advertising detection with both classifiers on FOOD and COS datasets are better than those of legal advertising detection with the same classifiers on the same datasets. The accuracy difference between illegal and legal advertising recognition with SVM classifier is comparatively smaller than that with Naïve Bayes classifier. Note that the ratio of legal instances versus illegal instances in the food dataset is 41.84:58.16, and the ratio in the cosmetic dataset is 48.03:51.97. Thirdly, in the first four feature settings, i.e., Unigram, Unigram+CILIN, Unigram+DOH, and Unigram+CILIN+DOH, the Naïve Bayes classifiers perform illegal advertising detection better than SVM classifiers in both datasets. In contrast, SVM classifiers achieve better legal advertising detection than Naïve Bayes classifiers. However, when the log relative frequency ratio is introduced, i.e., the Unigram+logRF feature setting, SVM classifier achieves the best performance in both illegal and legal advertising recognition on both datasets. The false alarm rates, a ratio of legal statements mis-recognized as illegal ones among all the legal statements, in food and cosmetic datasets are 0.083 and 0.166, respectively. The missing rates, a ratio of illegal statements mis-recognized as legal ones among all the illegal statements, in food and cosmetic datasets are 0.053 and 0.115, respectively. That illustrates the feasibility of the log relative frequency ratio and SVM classifiers.



| Classification Models →<br>Features ↓ Illegal vs. Legal → | Naïve Bayes    |                | SVM            |                |
|-----------------------------------------------------------|----------------|----------------|----------------|----------------|
|                                                           | Illegal        | Legal          | Illegal        | Legal          |
| Unigram                                                   | 92.592%        | 85.058%        | 89.463%        | 88.000%        |
| Unigram + CILIN                                           | 93.367%        | 83.851%        | 90.330%        | 88.889%        |
| Unigram + DOH                                             | 92.705%        | 84.994%        | 89.875%        | 89.106%        |
| Unigram + CILIN + DOH                                     | 93.421%        | 83.902%        | 90.159%        | 89.126%        |
| Unigram + logRF                                           | <b>94.317%</b> | <b>86.371%</b> | <b>94.696%</b> | <b>91.677%</b> |

Table 4.4 Individual accuracies of illegal and legal advertising recognition on food dataset

| Classification Models →<br>Features ↓ Illegal vs. Legal → | Naïve Bayes    |                | SVM            |                |
|-----------------------------------------------------------|----------------|----------------|----------------|----------------|
|                                                           | Illegal        | Legal          | Illegal        | Legal          |
| Unigram                                                   | 86.479%        | 77.632%        | 82.470%        | 82.357%        |
| Unigram + CILIN                                           | 86.812%        | 77.374%        | 83.287%        | 82.186%        |
| Unigram + DOH                                             | 86.944%        | 77.658%        | 83.375%        | 83.964%        |
| Unigram + CILIN + DOH                                     | 87.075%        | 77.431%        | 83.384%        | 83.260%        |
| Unigram + logRF                                           | <b>88.197%</b> | <b>83.060%</b> | <b>88.463%</b> | <b>83.413%</b> |

Table 4.5 Individual accuracies of illegal and legal advertising recognition on cosmetic dataset

## 4.4 Overstated Verb Phrase Mining

Effective identification of illegal advertising is a challenge for the authority and advertisers. Table 4.3 shows that almost all illegal advertising statements listed by DOH are verb phrases consisting of a transitive verb and an object noun. Thus, the usage of these verb phrases is a key criterion. To realize how illegal advertising uses verb phrases, illegal advertising verb phrases were mined from the illegal food and cosmetic datasets. The results can be used to extend the official list of illegal statements to improve the illegal advertising recognition processes performed by the authority and help advertisers prepare legal advertisements.

The first step of mining illegal advertising verb phrases is to obtain words that

present more frequently in the illegal datasets. To do this, the same formula of the log relative frequency ratio mentioned in Section 4.3 was used. If  $\log RF_{AB}(w^i)$  is a negative value, then  $w^i$  is more frequently used in illegal advertising. In these experiments, only the words with a log relative frequency lower than -0.1 and with appropriate POS tags were selected. The verb must be a transitive verb or nominalize verb, and the noun must be a common noun.

Next, each sentence in the datasets was examined to determine whether it contained a verb phrase consisting of a verb and a noun from the word list or not. Since a parser has not been used in this stage, and an object noun does not necessarily immediately follow its verb, a verb phrase is identified by the following criteria.

- (1) The verb should occur before the noun.
- (2) The distance between the verb and the noun should not exceed 3 words.

| Dataset | Illegal advertising verb phrases |                            |
|---------|----------------------------------|----------------------------|
|         | Transitive verb                  | Object noun                |
| FOOD    | 增強<br>(improve)                  | 體質<br>(physical condition) |
|         | 抑制<br>(inactivate)               | 細菌<br>(bacteria)           |
|         | 對抗<br>(fight)                    | 年齡<br>(aging)              |
|         | 分解<br>(decompose)                | 膽固醇<br>(cholesterol)       |
| COS     | 淨化<br>(purify)                   | 體質<br>(body system)        |
|         | 舒緩<br>(ease)                     | 疼痛<br>(pain)               |
|         | 供給<br>(provide)                  | 氧氣<br>(oxygen)             |
|         | 治療<br>(treat)                    | 面皰<br>(acne vulgaris)      |

Table 4.6 Example illegal verb phrases mined from the FOOD and COS datasets

The noun should be the head of the noun phrase where it presents. That is, the

noun should be the last word in the noun phrase. In Chinese, the head of a noun phrase is preceded by its adjectives and noun modifiers in most cases.

There are 979 verb phrases mined from the FOOD dataset and 2,302 verb phrases from the COS dataset. Some examples of these phrases are listed in Table 4.6.

The log relative frequency ratio can be used with a POS tagger to mine illegal verb phrases consisting of a transitive verb and an object noun. It can be seen that most verbs in the verb phrase lists are related to curative effects, and the objects are related to human body, nutrients and diet. Similar structure and properties can be seen in the sample illegal expressions provided by the government. Thus, the log relative frequency ratio is an effective method to construct illegal expression lists.

## **4.5 Building False Advertising Recognition System**

In this section, the construction of the False Advertising Recognition System, or **FAdR**, is described. The system was built with a web-based user interface and a backend recognizer programmed in Java. The models generated by WEKA through the training processes were directly used in the system. The structure is composed of the pre-processor, recognizer and explanation modules. Figure 4.2 shows the overall system architecture.

### **4.5.1 Pre-processing Module**

The classification models are sentence-based, so the main purpose of the Pre-processor module in the system is to detect sentence boundaries. Four types of punctuations, including period, colon, exclamation, and question mark, were used to segment a document into sentences. Line breaks were also regarded as a sentence boundary marker because many advertisements in Chinese put sentences in separate

lines without any punctuation. Sentences with less than three characters or more than 80 characters were ignored.

The word segmentation task was performed using the CKIP segmenter, which can be accessed through the TCP socket. Segmented data were represented by corresponding feature sets based on appropriate classification models and then converted to a format that the Recognizer can read as input.

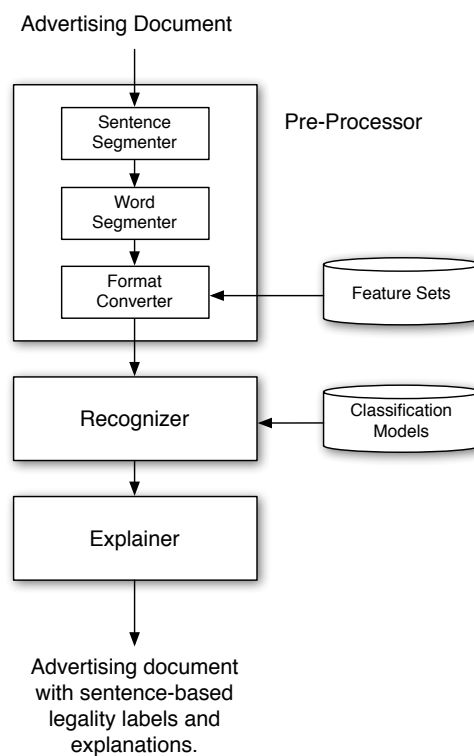


Figure 4.2 System architecture of **FAdR**

#### 4.5.2 Recognition Module

All processed sentences are sent from Pre-Processor to Recognizer for legality identification.

Since the training tasks were done in WEKA, the model files generated by WEKA can be used to implement Recognizer. Recognizer loads the pre-trained SVM models

for food and cosmetic advertising classification and then uses them for labeling the incoming sentences.

For the one-class models, the model files were pre-generated with different rejection rates from 0.4 to 0.9. When the user adjusts the threshold, Recognizer chooses the corresponding model to perform illegal sentences identification.

### 4.5.3 Explanation Module

To give users more information on possible reasons that specific advertising statements are considered illegal, Explainer uses the illegal verb phrase list, as discussed in Section 4.4, to extract inappropriate words from the input sentences. If the verb and object noun in a verb phrase from the list both occur in an illegal sentence, then the verb phrase will be shown with the recognition results in the user interface.

### 4.5.4 User Interface

Users can paste advertising texts in the text field or upload a document to the system. It usually takes less than 10 seconds on the server to process a document with 200 characters, so the system is suitable for quickly processing a large amount of data.

If users choose to use the one-class models, they can adjust the threshold value to fit different needs and receive useful results. Lowering the value can find as many potentially illegal sentences as possible, but more legal sentences can also be misjudged. Increasing the value can avoid wrongly labeling legal sentences as illegal, but more illegal sentences can be missed.

Figure 4.4 shows a system screenshot. The recognition results of a food advertisement with 11 sentences are demonstrated. Sentences labelled as illegal are highlighted in red. For illegal sentences, verb phrases possibly causing illegality are



listed in grey colour. The number of all sentences, the number of illegal sentences, and the final score are shown at the bottom. The correct score of an advertisement is defined as the number of correct sentences divided by total sentences in this advertisement. The sample advertisement used in Figure 4.4 and its English translation are shown in Figure 4.3.

日本茶第一品牌，全台首支融合三大天然色素的茶飲，可提升免疫力，消除壓力，增強體內抵抗力，增加體內抗體的形成。溫和不刺激，適合天天飲用。可降低自由基對細胞的過氧化傷害，強化人體免疫功能，健康好喝零負擔！

(The leading brand for Japanese tea. The first tea product combining three kinds of natural colourings in Taiwan. It can improve immunity, relieve stress, strengthen resistance to disease, and increase antibodies in your body. It is mild and not irritative. It is good for daily use. It can prevent body cells from being harmed by free radicals and strengthen immunity. It is healthy and tasty and brings no body burden.)

Figure 4.3 An illegal advertisement example



### Advertising Legality Identification

- Please enter the contents of your advertisement below.
- Sentences must be in Traditional Chinese and separated by full-width comma or period.

日本茶第一品牌，全台首支融合三大天然色素的茶飲，可提升免疫力，消除壓力，增強體內抵抗力，增加體內抗體的形成。溫和不刺激，適合天天飲用。可降低自由基對細胞的過氧化傷害，強化人體免疫功能，健康好喝零負擔！

Type:  Food  Cosmetic  One-class Shreshold:

|                 |                                                                                     |
|-----------------|-------------------------------------------------------------------------------------|
| 日本茶第一品牌         | ✔ Legal                                                                             |
| 全台首支融合三大天然色素的茶飲 | ✔ Legal                                                                             |
| 可提升免疫力          | ✘ Illegal                                                                           |
| 消除壓力            | ✘ Illegal<br>Possible cause(s):<br>消除, 壓力                                           |
| 增強體內抵抗力         | ✘ Illegal<br>Possible cause(s):<br>增強, 抵抗力                                          |
| 增加體內抗體的形成       | ✘ Illegal<br>Possible cause(s):<br>增加, 體內                                           |
| 溫和不刺激           | ✔ Legal                                                                             |
| 適合天天飲用          | ✔ Legal                                                                             |
| 可降低自由基對細胞的過氧化傷害 | ✘ Illegal<br>Possible cause(s):<br>降低, 傷害<br>降低, 細胞<br>降低, 自由基<br>氧化, 傷害<br>氧化, 自由基 |
| 強化人體免疫功能        | ✘ Illegal<br>Possible cause(s):<br>強化, 人體<br>強化, 功能                                 |
| 健康好喝零負擔         | ✔ Legal                                                                             |

Sentences examined: 11

Illegal sentences: 6

Score: 0.45

Figure 4.4 Screenshot of the FAdR recognition results

## Chapter 5 Verbal Irony



In verbal irony, the author/speaker intends to convey the opposite of what is literally expressed. To understand the actual meaning, the reader/listener needs to take context and background knowledge into account rather than just interpret the expression literally. Irony is a speech act used in special conditions and not a regular way to make daily communication. These characteristics make the retrieval, processing and analysis of ironic expressions a challenging task.

In this study, ironic messages are retrieved from the Plurk microblogging platform with a semi-automatic approach to construct an irony corpus. By utilizing this corpus, irony recognition approaches and an improved process of ironic message retrieval are proposed. The linguistic structure and clues of irony are also analyzed.

### 5.1 Irony and Sarcasm

The most widely accepted definition of verbal irony is the form of language usage where a speaker says something that is the opposite of what they mean (Gibbs and Colston, 2007). This represents the traditional semantic view of irony definition. From the pragmatic perspective, the social functions of irony can be humor, status elevation, aggression and emotional control, as discussed by Gibbs and Colston (2007). Merriam-Webster Dictionary provides a definition from both the semantic and pragmatic views, describing irony as “the use of words that mean the opposite of what you really think especially in order to be funny.”

As Wilson and Sperber (1992) and Gibbs and Colston (2007) have pointed out, the claim that irony always communicates the opposite of what is literally expressed has

some weaknesses. Actually, ironic understatements, quotations and interjections can also be observed in English (Wilson and Sperber, 1992).

This chapter will mainly explore ironic expressions in Traditional Chinese texts. Clarifying the definition of the Chinese equivalent of the English word “irony” is therefore necessary. In the official Chinese dictionary published by the Ministry of Education of Taiwan (MOE)<sup>7</sup>, “反諷” is basically defined in the same way as in English:

字面或言表與真正意念相反，藉以諷刺或增強語文力量。

*(The literal meaning is the opposite of the actual intentions. Used to be sarcastic or increase the intensity of words.)*

Based on that definition, only ironic expressions in which the opposite of the literal meaning is meant will be discussed in this chapter. That is, ironic expressions must contain one or more words/phrases that are not used literally. This means other kinds of verbal irony, such as ironic understatement described by Gibbs and Colston (2007) and hyperbole, rhetorical question and jocularity mentioned by Whalen et al. (2013), will not be covered. Situational and dramatic irony will also be excluded.

Sarcasm and irony has been studied by linguistics and cognitive scientists for years (Clark and Gerrig, 1984; Giora and Fein, 1999; Gibbs and Colston, 2007). Some studies have started focusing on the processing of sarcasm and irony recently, but it is still not clear whether sarcasm and irony differ significantly or represent the same concept. Some studies tried to distinguish the differences (Brown, 1980; Haiman, 1998), but recent studies on sarcasm and/or irony processing did not distinguish them (Davidov,

---

<sup>7</sup> <http://dict.revised.moe.edu.tw>

2010; Filatova, 2012).

In Merriam-Webster Dictionary, sarcasm is defined in a similar way as irony:



*The use of words that mean the opposite of what you really want to say especially in order to insult someone, to show irritation, or to be funny.*<sup>8</sup>

Besides the same description of irony, as shown above, the definition also mentions the purposes of sarcasm, including insult and irritation. Based on this interpretation, sarcasm is a speech act that is used to create negative effects. Gibbs and Colston (2007) also mention that “sarcastic irony is used to express a negative attitude as well as to insult or hurt to some degree” as well as “nonsarcastic irony can be used to express either negative or positive attitudes but in neither case is intended to hurt or insult any particular person“ (p. 93). These statements suggest that irony is not necessarily sarcastic. In MOE’s dictionary, the Chinese word 諷刺, which is the most common translation of the English word “sarcasm,” is described differently from 反諷:

*以隱微的方式嘲諷譏刺。*

*(to deride or satirize in an indirect way.)*

According to these descriptions, the words “irony” and “sarcasm” in English, as well as the words 反諷 and 諷刺 in Chinese, cannot be used interchangeably. In this study, the word “irony” will be used exclusively when referring to the expressions in which what is said is the opposite of what is meant.

---

<sup>8</sup> <http://www.merriam-webster.com/dictionary/sarcasm>



## 5.2 Irony Corpus Construction

This section introduces a bootstrapping methodology used to construct an irony corpus and mine irony patterns. While Lukin and Walker (2013) also used a bootstrapping method to improve sarcasm and nastiness classifiers, this study, in contrast, focuses on irony pattern mining and corpus construction. The procedure is shown in Figure 5.1.

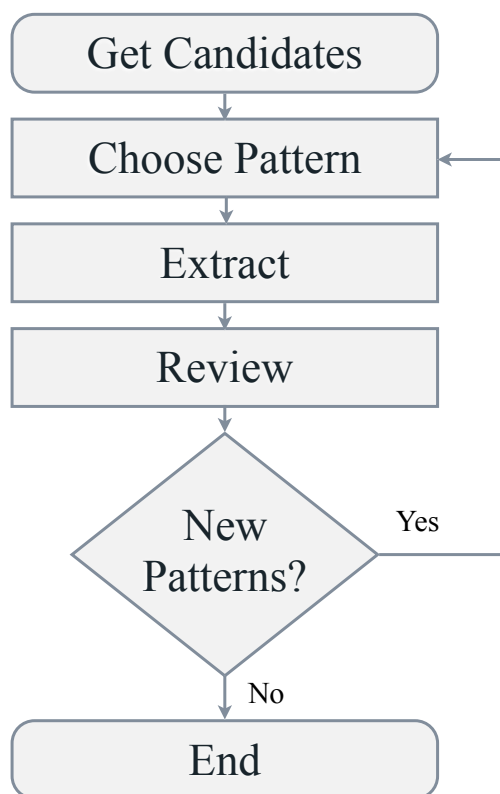
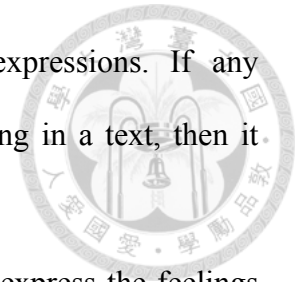


Figure 5.1 The bootstrapping procedure.

### 5.2.1 Finding Irony from an Emotion-Tagged Corpus

Under the definition of irony described in Section 5.1, texts annotated with information about actual meaning should be obtained, and the literal meanings of words

in the texts should be identified in order to collect ironic expressions. If any disagreement exists between the actual meaning and literal meaning in a text, then it may contain irony.



Nowadays, emoticons are used frequently on social media to express the feelings of the posters. These emotion icons specify a poster's actual meanings in some sense. Based on this idea, messages on Plurk, a microblogging platform described in Section 2.2, can be obtained for the irony processing purpose. Plurk lets users post messages limited to 140 characters and allows them to use graphical emoticons in their messages.

It is assumed that these emoticons can represent the poster's sentiments and therefore be regarded as sentiment labels of the messages. Among 35 emoticons, 23 are categorized into positive, while the other 12 are categorized into negative, as shown in Figure 2.2. Collected messages were generated from Jun 21, 2008 to Nov 7, 2009, and all of them are in Traditional Chinese.

On the other hand, the literal meanings of the posted messages need to be figured out. Among a variety of sentiment analysis algorithms (Liu, 2012), a lexicon-based approach is adopted for this study. The NTU Sentiment Dictionary, or NTUSD (Ku and Chen, 2007), was employed to determine the sentiment of a word. NTUSD contains 21,056 positive and 22,751 negative words. Most of these words are in Traditional Chinese.

There are some benefits to use microtexts like Plurk messages for this study of irony. The length of these messages is limited and usually shorter than a regular article, while they can still contain multiple sentences. This helps us exclude most irrelevant information but still capture discourse information.

### 5.2.2 Candidates Extraction

Possible irony messages were extracted from the Plurk corpus according to emoticons and NTUSD. Since the typical social function of irony is expressing negative meanings with positive words, as mentioned in Gibbs and Colston (2007), focus was directed on those messages with negative emoticons and positive words. A total of 3,178,372 messages are found containing at least one negative emoticon. Among them, 304,754 messages with at least one positive word are found and form an irony candidate dataset.

Discourse relation determines how two discourse units cohere to each other. Sentiment transition of two clausal arguments can be identified based on their discourse relation (Zhou et al., 2011; Wang et al., 2012; Huang et al., 2013). In the sentence “he is nice but not attractive,” positive opinion at the beginning is transformed to a negative one by the discourse connective “but.” In such a case, both the positive word “nice” and the negative phrase “not attractive” are used literally, and the sentence cannot be regarded as irony. For this reason, it is necessary to filter out messages containing such connectives. To do this, messages are removed only when the positive word occur earlier than the disjunctive word due to the grammatical structure of Chinese. The disjunctive words used in this step include 但, 但是, 可是, 只是, 不過 (all the above are equivalent to the English word but), 然而 (however), 卻 (comparatively), 可惜 (unfortunately), 偏偏 (contrarily), 反而 (oppositely), 倒是 (on the contrary). A total of 254,836 messages remain after this process.

### 5.2.3 Pattern Mining

Although irony can be used without any customary linguistic patterns, some ironic expressions do exhibit specific forms of language use. Colston and O'Brien (2000)



suggest that both irony and hyperbole create contrasts between expected and ensuing events. It is assumed that exaggerated expressions could be used with irony to strengthen the effects of the speech act. In the expression 我真是太幸運啦! (I am really and extremely lucky!), the adverbs *really* and *extremely* are used together to strengthen the ironic effect. Thus combinations of degree adverb phrases and a positive adjective are used as patterns in this study to find possible irony expressions automatically in the candidate dataset.

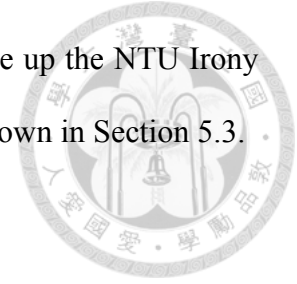
Not all degree adverbs in Chinese are used here because some of them are mostly used in formal texts and not frequently present in microblogs. The degree adverb phrases used here include the combinations of the adverbs “還” (*hái*), “也” (*yě*), “未免” (*wèimǎn*), “可” (*kě*) and “實在” (truly) and the degree adverbs “真” (really), “太” (extremely) and “非常” (very).

The following bootstrapping procedure is used to find more patterns.

- (1) Which patterns should be used is decided. At the very beginning of the bootstrapping procedure, the [degree adverb + positive adjective] pattern mentioned above is used.
- (2) Messages containing the patterns in step (1) are automatically retrieved from the candidates. NTUSD is used to determine sentiment polarity, and the CKIP parser is used to get parts of speech.
- (3) Messages retrieved in step (2) were reviewed by the annotator to decide which of them are actually ironic.
- (4) If the annotator finds new irony patterns in the reviewed messages, then the procedure starts again from step (1) and uses the patterns to repeat the process.

This process was repeated for four times. After the fourth iteration, no more new patterns were found by the annotator. Finally, 2,825 messages are found to have any of

the patterns, and 1,005 of them are confirmed to be ironic and make up the NTU Irony Corpus.<sup>9</sup> These patterns and examples of the ironic messages are shown in Section 5.3.



## 5.3 Irony Patterns

All the patterns mined by the approach mentioned in 5.2 are categorized into the following five groups.

### 5.3.1 Degree Adverbs + Positive Adjective

In this pattern, the following two components must exist:

- (a) Degree adverb phrase + positive adjective phrase
- (b) Negative context

The negative context can occur either before or after the component (a). For example, the following expression is used when someone has to wait for a long time to start ordering in a restaurant:

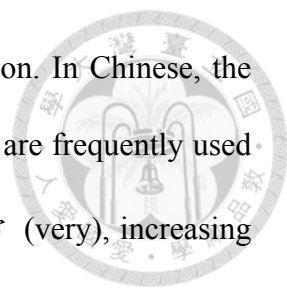
(s1) 點餐都要等半小時,服務還真是好阿

I have to wait for half an hour to order. The service is really good.

The underlined expression is the contextual information described in (b), and the double-underlined expression is the linguistic form described in (a). The adverbs 還 and 真 are combined in the second clause to form a degree adverb phrase and is followed by the adjective “good.” Although the positive word “good” is used, the speaker actually means the opposite. The first clause indicates why they think the

---

<sup>9</sup> The NTU Irony Corpus is available at [http://nlg.csie.ntu.edu.tw/nlpresource/irony\\_corpus/](http://nlg.csie.ntu.edu.tw/nlpresource/irony_corpus/)



service is not good and therefore provides the contextual information. In Chinese, the adverbs 還 (hái), 也 (yě), 未免 (wèimǎn), 可 (kě) and 實在 (truly) are frequently used to modify the degree adverbs 真 (really), 太 (extremely), and 非常 (very), increasing the effect of an ironic expression. For example, the adverb phrases 也大, 未免大, 可真, and 還真 can be seen frequently in the corpus.

The positive adjective or adjective phrase used in this pattern is the expression that cannot be interpreted literally. The most common words for this include 好 (good), 棒 (excellent), 有趣 (interesting), 幸運 (lucky) and 美好 (wonderful).

The degree adverbs mentioned above have high intensity in meaning and are used to modify positive adverbs. Even though they seem to intensify the positive attributes described by the adjective, the speaker's highly negative feelings are actually expressed. This hyperbolic use of adverbs can be frequently seen in our irony corpus.

### 5.3.2 The Use of Positive Adjective with High Intensity

In this pattern, the following two components must exist:

- (a) Positive adjective with high intensity
- (b) Negative context

Specific positive adjectives with high intensity are used to form ironic expressions with or without other rhetorical elements. Since the context is negative, the positive adjective is used to express non-literal meanings. This kind of adjectives found in the corpus includes 偉大 (great), 了不起 (remarkable) and 天才 (genius). For example, the word “great” is used in the following message:

(s2) 我的 plurk「又」發生不明錯誤了...這真是這世紀最偉大的發明啊

My Plurk account encountered an unknown error “again”... This is indeed the greatest invention in the century.



### 5.3.3 The Use of Positive Noun with High Intensity

In this pattern, the following two components must exist:

- (a) Positive noun with high intensity
- (b) Negative context

Specific nouns that represent highly positive meanings are also used to express irony. These nouns include 巨星 (superstar), 大禮 (big gift) and 境界 (wonderful state). When they are used with negative context, an ironic expression is formed. An example is listed below:

(s3) 中秋節收到的大禮是.....長了一堆肉

The big gift I received in the Mid Autumn Festival was..... a lot of fat in my body.

### 5.3.4 The Use of 很好 (very good)

In this pattern, the following two components must exist:

- (a) Sentence boundary + 很好 + punctuation
- (b) Negative context

A sentence boundary occurs before the word 很好 (very good) because there is no subject. Multiple punctuations, especially exclamation marks and ellipses, can be used after 很好 to increase the intensity. In the following example, exclamation marks are

used:



(s4) 感冒... 很好!! 我的假期飛了

I caught a cold... Very good!! My vacation is gone.

Sometimes this pattern is followed by an exclamation word, such as 啊 (a), 呀 (ya) and 嘛 (ma). These exclamations, like punctuations, can help strengthen the level of the speaker's feelings. In our irony corpus, this pattern is used in 46.36% of all ironic messages. Obviously, this is a common way for Plurk users to express their negative feelings with an ironic expression.

### 5.3.5 可以再...一點 (“It's okay for someone/something to be worse”)

In this pattern, the following expression must exist:

可以再 + negative adjective + 一點 (It is okay to be more + negative adjective)

This pattern literally states that it is okay for something to become worse. This is a commonly used pattern to express irony in Chinese. In most cases, even when no proper contextual information is present, the listener can tell the literal meaning isn't meant in this pattern because it violate most people's inclinations. An example is shown below.

(s5) 零下十一度...你可以再冷一點

It's -11°C... It is fine for the weather to be colder

The expression 沒關係 (“it is okay” or “I am fine with that”) is frequently added after the pattern. For example:



(s6) 網路可以再慢一點啊..沒關係呀..

It is okay if the network becomes slower .. I am fine with that...

In examples (s5) and (s6), contextual information is provided. However, contextual information is missing in many utterances containing this pattern. In this case, only the adjective can be used to determine whether it is ironic or not. Since the same adjective can be either positive or negative based on the properties of the subject it modifies, this kind of irony can be incorrectly recognized in the irony recognition task.

The patterns in 4.3.4 and 4.3.5 are mainly based on their linguistic forms and frequently used in ironic expressions. It can be seen that these patterns are more static than the others, and they will be called *customary patterns* in this chapter. On the other hand, the patterns in 4.3.1, 4.3.2 and 4.3.3 are called *non-customary patterns*.

## 5.4 Linguistic Structure of Irony

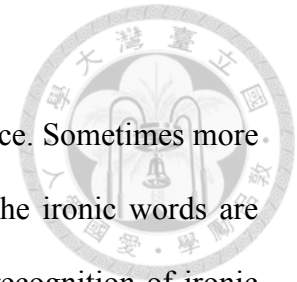
The linguistic structure of verbal irony may not be obvious, and no comprehensive theories have been proposed by linguists to describe it. We believe that the first step to understand the structure of irony is to identify the elements constituting irony. Based on the patterns found in Section 5.3 and the usage observed in the corpus, three elements can be found to be used to convey irony.

### 5.4.1 Ironic Word/Phrase

As described, the literal meaning of an ironic word or phrase is opposite to the actual meaning. An ironic word/phrase is necessary to separate irony from regular utterances. If the ironic word of an utterance is reverted, the speaker's actual sentiment

or intention is reconstructed.

However, it is not easy to identify the ironic word in an utterance. Sometimes more than one word can be an ironic word. In our corpus, 94.93% of the ironic words are adjectives, while others are used as adverbs, verbs or nouns. The recognition of ironic word/phrase is a challenging task. An ironic word recognition approach is proposed in Section 5.6.



#### 5.4.2 Contextual Information

Contextual information is usually provided as part of an ironic utterance to help convey irony. For example, the underlined sentence in the following utterance is crucial for irony interpretation:

(s7) 我掛彩了,真是太好運了

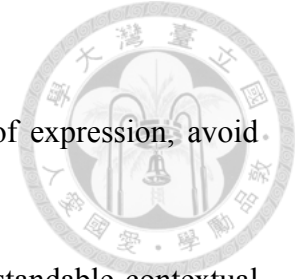
I was injured. I was really lucky.

Without the first sentence, it would be hard to tell if “lucky” is actually meant. Although a speaker can use ironic words/phrases without providing contextual information, this can be an ineffective way to communicate the actual meanings of irony. According to the cooperative principle proposed by Grice (1975), the speaker must give enough information in order to enable successful communication and implicatures. The four maxims of the cooperative principle include:

- (1) **Maxim of Quantity:** The speaker should make their contribution as informative as is required. Do not make the contribution more informative than is required.
- (2) **Maxim of Quality:** The speaker should not say what they believe to be false, and should not say that for which they lack adequate evidence.

(3) **Maxim of Relation:** The speaker should be relevant.

(4) **Maxim of Manner:** The speaker should avoid obscurity of expression, avoid ambiguity, be brief and be orderly.



Based on these maxims, enough, correct, relevant, and understandable contextual information should be provided with ironic expressions. However, the speaker sometimes assumes that the listener already knows about the conditions where the irony takes place and has the required background knowledge, and thus the contextual information is hidden in the ironic utterance.

Four types of context can be used to interpret irony:

(1) Linguistic context

The linguistic context refers to the words that are expressed before and/or after the irony words in a sentence or discourse. It is easier to obtain and analyze than the other three types of context.

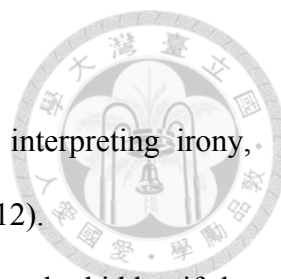
(2) Physical context

Physical context refers to what is actually present and/or happening in the environment or circumstance where the conversation is taking place. It is also related to the timing. In online conversations, participants are not usually in the same location, but they can be aware of the same ongoing events and situations. It is not necessary for the speaker to provide physical context information if they assume the objects or situations is noticeable to the listeners.

(3) Epistemic context

The background knowledge shared by the participants in a conversation can also be used to interpret irony. This type of context does not change over time. For example, people know rocks are hard, so they can understand the expression *the bed is as soft as a rock* is not literal.





#### (4) Social context

Social relationship can be important for expressing and interpreting irony, especially in online messages. An example is provided in (s12).

At least one type of contextual information must exist, but it can be hidden if the speaker thinks the listener is already aware of it. Physical, epistemic and social context can be hidden, while linguistic contextual information must be present.

### 5.4.3 Rhetoric

As shown in Section 5.3, degree adverbs, punctuations and exclamations can be used to convey irony. Some of them can even be repeated to intensify the effects. These elements increase contradiction and strengthen the degree of negative opinions. Unlike ironic words and context, rhetoric elements are not necessary to convey irony.

Liebrecht et al. (2013) call the words used to strengthen evaluative utterances *intensifiers*. In their experiments, non-hyperbolic sarcastic messages often contain an explicit marker on Twitter. They argue that sarcasm is often signaled by hyperbolic words, including intensifiers and exclamations, and sarcastic utterances with hyperbolic words are easier to identify by listeners/readers than sarcastic utterances without hyperbolic words. It can be seen that adverbs, adjectives, punctuations and exclamations with high intensity observed in our irony patterns have very similar effects.

### 5.4.4 Corpus Labeling

To increase the usefulness of the corpus, ironic element tags are added to each message. An example is shown in Figure 5.2.

```
<context sentiment="neg">才剛買的書，竟然掉頁了，</context>這  
品質<rhetic>也太</rhetic><ironic sentiment="pos">好</  
ironic>了<rhetic>吧</rhetic>。
```

English translation:

```
<context sentiment="neg">The book I just bought has fallen  
apart.</context> The quality is <rhetic>just extremely</rhetic>  
<ironic sentiment="pos">good</ironic>le<rhetic>ba</  
rhetic>.
```

Figure 5.2 An example message with ironic element tags

As can be seen in the example message, 好 (good) is the word that is used in the opposite way, so it is marked with the ironic word/phrase label `<ironic>`. The preceding sentence states what actually happened, and is marked with the label `<context>`. The message poster didn't just use the positive word "good," but also modified it with the degree adverb phrase 也太 and used the exclamation 吧 (*ba*, a sentence-final particle) at the end of the sentence. These two elements are marked with the `<rhetic>` label. The sentiment polarity marks of the ironic word and contextual information, shown as either `pos` or `neg`, are also added.

## 5.5 Clues to Irony

Ironic utterances are indirect and ambiguous, and the author/speaker thus risks being misunderstood by the reader/listener. That is, the reader/listener may regard the literal meaning as the actual meaning. To accomplish the speech act of irony, the author/speaker must ensure that the reader/listener has or will receive adequate information to correctly interpret the ironic utterance. As mentioned in Section 5.4.2, the author/speaker normally obeys the cooperative principle as suggested by Grice (1975) and tries to provide necessary contextual information. Gibbs and Colston (2007)

suggest that intentionally violating Gricean Maxims would result in an ironic interpretation (pp. 126-127). The superficial violation of the maxims therefore creates the clues to irony that can be used by the reader/listener to avoid misinterpretation.

The clues to be discussed below are lexical or non-lexical devices used by the author/speaker to convey irony. By examining these clues in irony, the interpretation of irony, that is, how irony is indicated by the author/speaker and finally successfully perceived as irony by the reader/listener, can be further explored. An irony clue can be unveiled in one or more irony elements described in Section 5.4 or through the existing knowledge and understanding of the real circumstances.

Booth (1974) suggests that five kinds of clues can be provided by the author: (1) straightforward warnings in the author's own voice, (2) known error proclaimed, (3) conflicts of facts within the work, and (4) clashes of style. In face-to-face communication, the speaker can make use of paralinguistic clues including intonation, loudness, gesture and facial expression. In this study, however, written materials are used exclusively for ironic expression collection and analysis, making obtaining paralinguistic components impossible. This doesn't mean that non-lexical elements cannot be used as clues in online messages. The special usage of letters, characters, symbols and emoticons are found to be related to irony. The three different kinds of clues to irony are found in the NTU Irony Corpus:

(1) Customary ironic expression

Some linguistic patterns, such as 可以再...一點 mentioned in Section 5.3.5, are used regularly to convey irony. They can contain the ironic word/phrase and some accompanying lexical elements. When such a pattern is present, there is a high probability of irony. For example:

(s8) 辦公室咖啡可以再難喝一點。

It's fine if the office's coffee tastes even more awful.



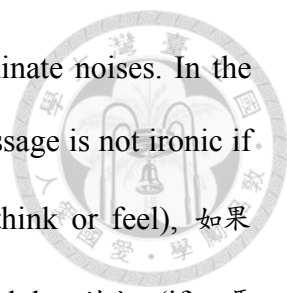
Table 5.1 shows how likely a message is ironic when a specific pattern is present. In 1,000 messages containing the pattern [可以再 + any word(s) + 一點] randomly chosen from the Plurk dataset, there are 78.3% confirmed to be ironic. As for the pattern [#很好 + punctuation], where the symbol # indicates the beginning of a message, 60.5% of them are ironic. This suggests that 可以再...一點 is a more powerful customary irony pattern than #很好。

| Pattern           | Irony Likelihood |
|-------------------|------------------|
| 可以再...一點          | 78.30%           |
| #很好 + punctuation | 60.50%           |

Table 5.1 Likelihoods of irony for random 1,000 messages with specific patterns

Some linguistic patterns originated from a specific linguistic community, and their uses are only popular within a group of people. For example, 丁丁是個人才, 很好很強大, 我開始相信你了 and 記者快來抄 emerged on blogging platforms and online forums and later became popular throughout the Internet.

It is possible that the reader identify this kind of irony even if there is no contextual information is present in the author's message because the customary pattern itself is a clue to irony. The message (s8) shown above does not include any other information except for the sentences containing the irony patterns. The likelihood of 可以再...一點 is calculated with a dataset that



contains the pattern but has not been preprocessed to eliminate noises. In the 1,000 message containing this pattern, it is found that a message is not ironic if the pattern is preceded by the word 希望 (hope), 覺得 (think or feel), 如果 (if), 以為 (think), 應該 (should) or 還 (still) or followed by 的話 (if), 嗎 (ma, question marker), 了 (le, perfective aspect marker) or 就好了 (if..., it would be good.). Most of these words are used to indicate an expectation, opinion or question. The following is an example:

(s9) 午休時間可以再長一點嗎!

Can the length of the noon break be longer!

To understand how these words affect the usage of the irony pattern 可以再...一點, other 1,000 messages [可以再 + any word(s) + 一點] are randomly chosen from the Plurk dataset. If the messages with those words are removed in advance, the likelihood of irony increases to 88.20%, indicating that the pattern is a reliable clue to irony.

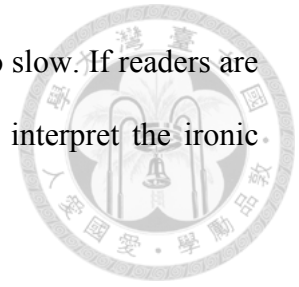
## (2) Irrational statement

When a message contains an “error” or contradiction, the author either makes a mistake or is intentionally being ironic. If the reader does not think the author is careless, then an ironic interpretation is made. The error can occur in physical, epistemic and social contexts:

### a. Error in physical context

In the following example, the author states that he or she is very

“patient” while the website he or she is using is too slow. If readers are aware of the status of the website, then they can interpret the ironic word “patient” correctly.



(s10) 樂 X 的網站~~真的可以再慢一點~~我非常的有耐心~~

It is fine that the website is so slow~~I am very patient~~

b. Error in epistemic context

In the following example, the author and the reader share the knowledge of computer viruses, and thus accomplish a successful irony interpretation.

(s11) 電腦半年中毒三次還真棒

It is wonderful that the computer has been infected with viruses three times in six months.

c. Error in social context

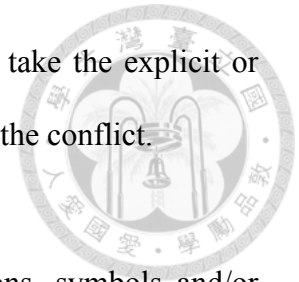
While seeing the boss in a holiday is usually not a good thing, whether or not the following message is ironic still depends on the author's current actual relationship with the boss:

(s12) 假日還要見老闆..真是太幸福了

I have to see my boss in the holiday..I really have a happy life.

In order to perceive these expressions as irony, it is necessary for the

reader/listener to identify the ironic word/phrase and also take the explicit or implicit contextual information into account in order to see the conflict.



### (3) Rhetoric

Special uses of words, letters, characters, punctuations, symbols and/or textual or image emoticons can be found in our irony corpus. Rhetoric and other non-lexical elements do not cause opposite meanings, but they can be used as irony indicators and to intensify the effects of irony. When an expression is unusual or overstated in terms of style, it draws the reader's attention and makes them more likely to interpret the expression correctly.

Duplication, hyperbole and intensifiers can be found in our irony corpus:

#### (1) Duplication

Words, punctuations and other symbols can be repeated to help convey irony.

(s13) 今天真是太太太幸運了~在台九線上開車，大卡車迎面而來...

I was very very very lucky today~ When I was driving on the highway, a large truck rushed toward me...

#### (2) Hyperbole

Intentional emphasis on or exaggeration of something can make a clue to irony.

(s14) 亞洲第一名 台灣的驕傲!!

The best in Asia. The pride of Taiwan!!

(s15) 這真是這世紀最偉大的發明啊

This is indeed the greatest invention in the century.



### (3) Intensifier

Some degree adverbs and complements can be used as intensifiers.

(s16) 今天真是運氣好到爆耶!!!

I am crazy lucky today!!!

Any words or phrase that are used as rhetoric clues will be labeled as rhetoric elements in the irony corpus.

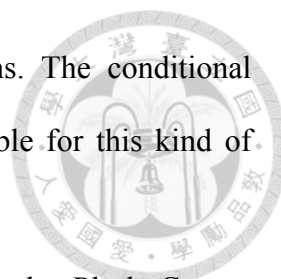
## 5.6 Irony Recognition

As described in 5.4.1, the ironic word in an ironic utterance is the word that should be interpreted in the opposite way. The correct meaning of an ironic expression can be constructed more easily if the ironic word is recognized. In this section, ironic word and ironic message recognition algorithms are proposed.

### 5.6.1 CRF-based Algorithms

Since each of the irony patterns found in Section 5.3 is an ordered arrangement of elements represented by specific words and parts of speech, it is assumed that irony is a sequential process where states transition from one to another. To identify the structure of irony and the presence of the ironic word/phrase, the irony elements proposed in Section 5.4 can be used as hidden states in a hidden Markov process. Similarly, the





words and their parts of speech can be regarded as observations. The conditional random field (CRF) algorithm (Lafferty et al., 2001) is thus suitable for this kind of sequential modeling. The following configurations are adopted:

- (1) Datasets are extracted from the NTU Irony Corpus or the Plurk Corpus depending on the requirement of each experiment.
- (2) The CKIP word segmentation system is adopted for Chinese word segmentation and POS tagging.
- (3) The irony element labels in the NTU Irony Corpus are used to determine the state of each word in a message. The four states are ironic word, rhetoric, context and unlabeled word, represented by 1, 2, 3, 0, respectively.
- (4) To implement the CRF algorithm, the CRF++<sup>10</sup> tool is adopted. The feature template specified in CRF++ is shown in Figure 5.3.

For example, the message “很好!在農場裡貧富差距很大, 有錢人都沒得偷” is represented by:

Words: {很, 好, !, 在, 農, 場, 裡, 貧, 富, 差, 距, 很, 大, , , 有, 錢, 人, 都, 沒, 得, 偷}

POS: { Dfa, VH, EXC, P, Nc, Ncd, Na, Na, Dfa, VH, COMM, Na, D, D, D, VD}

States: {2, 1, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0}

### 5.6.2 Irony Element Identification

In the NTU Irony Corpus, irony elements including ironic word/phrase, rhetoric and context are labeled. Correct recognition of these elements can improve the analysis and processing of ironic expressions. For example, if the ironic word is found, the speaker’s actual meaning can be obtained by reversing the meaning of the word/phrase.

---

<sup>10</sup> <https://taku910.github.io/crfpp/>



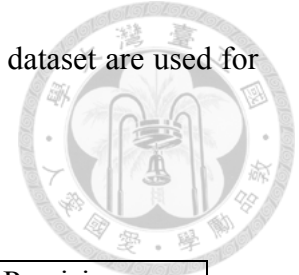
```
U00:%x[-2,0]
U01:%x[-1,0]
U02:%x[0,0]
U03:%x[1,0]
U04:%x[2,0]
U05:%x[-1,0]/%x[0,0]
U06:%x[0,0]/%x[1,0]
U07:%x[-2,1]
U08:%x[-1,1]
U09:%x[0,1]
U10:%x[1,1]
U11:%x[2,1]
U12:%x[-2,1]/%x[-1,1]
U13:%x[-1,1]/%x[0,1]
U14:%x[0,1]/%x[1,1]
U15:%x[1,1]/%x[2,1]
U16:%x[-2,1]/%x[-1,1]/%x[0,1]
U17:%x[-1,1]/%x[0,1]/%x[1,1]
U18:%x[0,1]/%x[1,1]/%x[2,1]
```

Figure 5.3 Feature template used in CRF++

Among the irony patterns mentioned in Section 5.3, the use of the combinations of adverbs is one of the most unidentifiable ironic expressions. Since “真” (*zhen*) is the most common character used in these adverb combinations, a dataset that contains all messages with the character *zhen* from the NTU Irony Corpus is used to perform the experiment of irony element identification. The experiment is described below:

- (1) All messages containing the character *zhen* are retrieved to form the dataset regardless of the usage of the character. There are totally 202 messages in this dataset.
- (2) All messages are preprocessed using CKIP for word segmentation and POS tagging.
- (3) The state of each word is defined by the irony element labeled in the NTU Irony Corpus.
- (4) The CRF models and features mentioned in 5.6.1 are adopted for the training

and testing. Two-thirds of all the messages from the *zhen* dataset are used for training, while the rest are used for testing.



| Target \ Measure                    | Recall | Precision |
|-------------------------------------|--------|-----------|
| Ironic word                         | 87.03% | 75.23%    |
| Any Rhetoric Element                | 63.50% | 72.57%    |
| <i>Zhen</i> as rhetoric element     | 88.24% | 86.54%    |
| <i>Zhen</i> as non-rhetoric element | 75.00% | 77.78%    |
| Context                             | 80.95% | 77.40%    |

Table 5.2 Experimental results of irony element identification

The results are shown in Table 5.2. The recall for ironic word identification is 87.03% and the precision is 75.23%. As for the identification of any kinds of rhetoric element, the recall is 63.5% while the precision is 72.57%.

When the performance of the identification of the rhetoric element *zhen* is specifically calculated, that is, when all rhetoric elements other than *zhen* are not taken into account, the recall and precision are 88.24% and 86.54%, respectively. On the other hand, *zhen* can also be used for non-ironic purposes in some messages. To understand if this kind of *zhen* can be correctly tagged as a regular word rather than an irony related use, the identification of the instances of *zhen* that are not used as rhetoric elements is also evaluated. The recall and precision are 75.00% and 77.78%, respectively.

### 5.6.3 Ironic Message Identification

A similar CRF-based algorithm is also used to identify ironic messages in a subset containing specific adverb combinations. The purpose of this task is to improve the precision of the bootstrapping procedure shown in Section 5.2 and thus reduce human effort. All of the messages with 還真, 可真, 也真, 也太 and/or 真是太 are collected from the NTU Irony Corpus for training, and 3,000 messages with these adverbs are

randomly collected from the Plurk corpus to form a test set. The experiment is performed using the following steps:

- (1) The CRF model is trained using the 139 messages containing the adverb combinations mentioned above. The same features shown in 5.6.2 are used. The test process results in 773 messages that are identified as irony.
- (2) If the adverb combination is not followed by a negative adjective, then the messages containing it is removed. 383 messages remain after this step.
- (3) The messages with positive context are removed based on NTUSD. The clauses without the adverb combination are assumed to be context. 197 messages are left after this step.
- (4) The messages with at least one disjunctive are removed. 170 messages after this step.
- (5) Finally, 47 messages are identified and confirmed to be ironic by human. The precision is 27.65%.

The above approach integrated with the CRF-based algorithm improves the bootstrapping approach shown in Section 5.2. As a result, human effort is reduced. Some examples of the ironic messages found by this approach are listed below:

- (1) 大家都在拼滿壘不得分,這場比賽還真是夠君子的(大噓
- (2) 飛機又墜毀了...台灣的國防還真是安全啊
- (3) 誠品的效率還真是好.....一年半前訂 Ulrich Schnauss 的專輯,終於到貨了.....

The factors that cause the situations where some ironic messages are not captured by this approach or some non-ironic messages are identified as irony by the model are analyzed and listed below.

- (1) The quality and appropriateness of the positive and negative word list in NTUSD

- (2) Complicated or unstructured grammatical structure
- (3) Multiple topics or referents shown in a message
- (4) Unclear sentiment polarity that depends on background knowledge or facts that cannot be easily identified in the message



More effective approaches may need to be explored and proposed in the future in order to more accurately identify these factors and improve irony recognition performance.

## Chapter 6 Conclusion



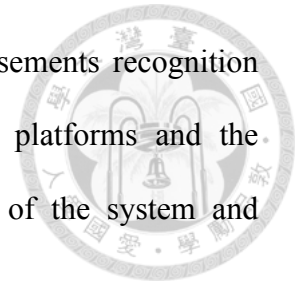
This study explored three challenging topics in natural language processing and linguistics—namely emotion modeling, advertising legality identification and irony analysis—based on microtext materials. The intent of the microtext author and/or readers hidden behind a short text were discovered and analyzed.

In this study, emotion generation models and transition models are proposed. Both writers' and readers' perspectives on emotion modeling are presented with regard to emotion recognition in microblog posts. Since the Plurk microblogging platform is also a social networking system, social relation, user behavior and relevance degree features were used along with textual features to build classifiers. As a result, the former two types of features proved useful to achieving better emotion-recognition performance. This study also shows that predicting emotion from readers' perspective is more challenging than from writers' perspective.

The results of the emotion analysis suggest that a reader model should be treated differently from a writer model. The same bigram or word can have different effects on writers' and readers' emotional expression. For example, greetings can cause a positive reader response even if the writer uses a negative emoticon. These findings suggest that reader emotions need to be further studied in the future.

As for the advertising legality recognition experiments, three subtopics were covered: illegal advertising statement recognition, illegal advertising verb phrase mining and the construction of an automatic false online advertisement recognition system. To identify the legality of a short advertising statements, log relative frequency ratio was found to be useful when used as weight of textual features. Log relative frequency ratio were also used for illegal advertising verb phrase mining. By combining

the techniques and results of the above two tasks, a false advertisements recognition system was built. Internet users, advertisers, online advertising platforms and the authorities all can benefit from the efficiency and convenience of the system and minimize the damage caused by false online adverteng.



In this study, the NTU Irony Corpus, which contains more than 1,000 Chinese ironic expressions in the form of microtexts, was constructed based on linguistic forms and sentiment classification. To the best of our knowledge, this is the first Chinese irony corpus that is annotated with irony element labels. To build this corpus, a bootstrapping procedure was used to reduce human effort. In addition, the linguistic structure of irony was also explored. Ironic word and phrases, contextual information, and rhetoric were found to comprise an ironic expression.

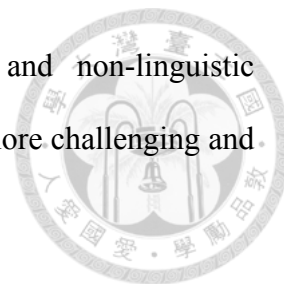
However, more irony patterns and phenomena have yet been discovered and should be examined in future studies. For instance, the following types of irony are not included in the NTU Irony Corpus:

- (1) Literally negative expressions that are actually positive
- (2) Ironic understatements
- (3) Situational irony

Up to now, irony in the Chinese language has not been thoroughly investigated by either linguists or computer scientists. In order to improve the performance of irony detection, more studies on the linguistic aspects of irony are needed in the future. The diversity of irony corpora should also be increased by including different types and linguistic patterns of irony.

As shown in the data observed in this study, other types of irony or sarcasm—e.g., ironic understatements, situational irony and the irony in which positive meanings are represented by negative literal meanings—cannot be seen frequently in microtexts, and

the retrieval of these ironic texts requires more contextual and non-linguistic information. The identification of these kinds of ironic use is even more challenging and needs to be further explored in the future.

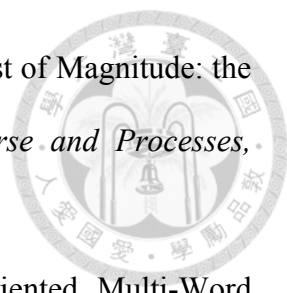


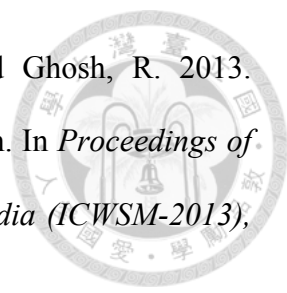


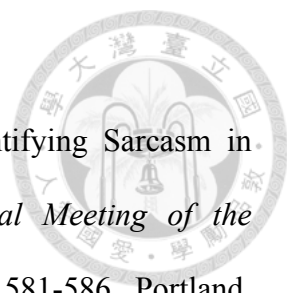
## REFERENCE

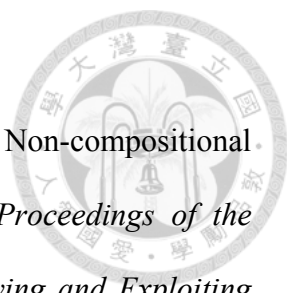


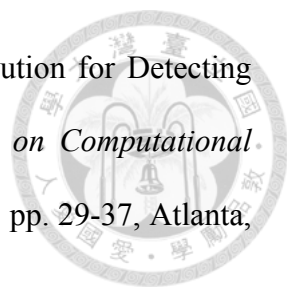
- Aman, S. and Szpakowicz, S. 2007. Identifying Expressions of Emotion in Text. In *Proceedings of 10th International Conference on Text, Speech and Dialogue. Lecture Notes in Computer Science 4629*, pp. 196-205.
- Becker, C., Kopp, S. and Wachsmuth, I. 2004. Simulating the Emotion Dynamics of a Multimodal Conversational Agent. In *Proceedings of Tutorial and Research Workshop on Affective Dialogue Systems*, pp. 154-165.
- Bernhaupt, R., Boldt, A., Mirlacher, T., Wilfinger, D. and Tscheligi, M. 2007. Using Emotion in Games: Emotional Flowers. In *Proceedings of the International Conference on Advances in Computer Entertainment Technology*, 41-48.
- CFIA. 2010. Advertising Requirements. Canadian Food Inspection Agency. Available at <http://www.inspection.gc.ca/english/fssa/labeti/advpube.shtml>.
- Chang, C.C. and Lin C.J. 2001. LIBSVM: a Library for Support Vector Machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Chao, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. University of California Press.
- Chen, H.H., Lin, C.C. and Lin, W.C. 2002. Building a Chinese-English WordNet for Translingual Applications. *ACM Transactions on Asian Language Information Processing*, 1(2): 103-122.
- Chen K.J. and Hsieh Y.M. 2004. Chinese Treebanks and Grammar Extraction. In *Proceedings of International Joint Conference on Natural Language Processing*, pp. 560-565.
- Cheng, H and Cantú-Paz, E. 2010. Personalized click prediction in sponsored search. In *Third ACM International Conference on Web Search and Data Mining (WSDM 2010)*, pp. 351-359, New York, USA.

- 
- Colston, H.L. and O'Brien, J. 2000. Contrast of Kind Versus Contrast of Magnitude: the Pragmatic Accomplishments of Irony and Hyperbole. *Discourse and Processes*, 30(3):179-199.
- Damerau, Fred J. 1993. Generating and Evaluating Domain-Oriented Multi-Word Terms from Text. *Information Processing and Management*, 29:433-477.
- Davidov, D., Tsur, O. and Rappoport, A. 2010. Semi-Supervised Recognition of Sarcastic Sentences in Twitter and Amazon, In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL-2010)*, pp. 107-116, Uppsala, Sweden.
- Dent, K. and Paul S. 2011. Through the Twitter Glass: Detecting Questions in Micro-Text. In *Workshop on Analyzing Microtext at the 25th AAAI Conference on Artificial Intelligence*.
- Derczynski L., Ritter A., Clark S. and Bontcheva K. 2013. Twitter Part-of-Speech Tagging for All: Overcoming Sparse and Noisy Data. In *Proceedings of Recent Advances in Natural Language Processing*, pp. 198-206. Hissar, Bulgaria.
- DOH (2009). Legal and Illegal Advertising Statements for Cosmetic Regulations. Department of Health of Taiwan, Available at <http://www.doh.gov.tw/ufile/doc/0980305527.pdf>.
- Edelman, B., Ostrovsky, M. and Schwarz, M. 2007. Internet Advertising and the Generalized Second Price Auction: Selling Billions of Dollars Worth of Keywords. *American Economic Review*, 97(1):242-259.
- Ellen, J. 2011. All about Microtext: A Working Definition and a Survey of Current Microtext Research within Artificial Intelligence and Natural Language Processing. In *Proceedings of the Third International Conference on Agents and Artificial Intelligence*.


- 
- Fei, G., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M. and Ghosh, R. 2013. Exploiting Burstiness in Reviews for Review Spammer Detection. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM-2013)*, pp. 175-184.
- Filatova, E. 2012. Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, pp. 392-298, Istanbul, Turkey.
- FTC. 2000. Advertising and Marketing on the Internet: Rules of the Road, Bureau of Consumer Protection. Federal Trade Commission, September 2000, Available at <http://business.ftc.gov/sites/default/files/pdf/bus28-advertising-and-marketing-internet-rules-road.pdf>
- Gabrilovich, E., Josifovski, V. and Pang, B. 2008. Introduction to Computational Advertising. *Tutorial Abstracts of ACL-08: HLT*, p. 1.
- Gabrilovich, E., Josifovski, V. and Pang, B. 2009. Introduction to Computational Advertising. IJCAI 2009 Tutorial, [http://research.yahoo.com/tutorials/ijcai09\\_compadv/](http://research.yahoo.com/tutorials/ijcai09_compadv/)
- Ghosh, A., McAfee, P., Papineni, K. and Vassilvitskii, S. 2009. Bidding for Representative Allocations for Display Advertising. CoRR, abs/0910-0880, 2009.
- Gibbs, R.W. and Colston, H.L. 2007. *Irony in Language and Thought*. Lawrence Erlbaum Associates, New York.
- Giora, R. and Fein, O. 1999. Irony: Context and Salience. *Metaphor and Symbol*, 14:241-257.
- Go, A., Huang, L. and Bhayani, R. 2009. *Twitter Sentiment Classification Using Distant Supervision*. CS224N Project Report, Stanford University, Stanford, CA.
- Gokhman, S., Hancock, J., Prabhu, P., Ott, M. and Cardie, C. 2012. In Search of a Gold Standard in Studies of Deception. In *Proceedings of the EACL 2012 Workshop on*

- 
- Computational Approaches to Deception Detection*, pp. 23-30.
- González-Ibáñez, R., Muresan, S. and Wacholder, N. 2011. Identifying Sarcasm in Twitter: A Closer Look. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Short Papers*, pp. 581-586, Portland, Oregon, USA.
- Grice, H. P. 1975. Logic and Conversation. In *P. Cole and J. J. Morgan, eds. Syntax and Semantics, 3: Speech Acts*. New York: Academic Press.
- Gupta, N., Gilbert, M. and Fabrizio G.D. 2010. Emotion Detection in Email Customer Care. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 10-16.
- Hempstalk, K., Frank, E. and Witten, I.H. 2008. One-Class Classification by Combining Density and Class Probability Estimation. In *Proceedings of the 12th European Conference on Principles and Practice of Knowledge Discovery in Databases and 19th European Conference on Machine Learning*, pp. 505-519.
- Huang, H.C., Lin, M.S. and Chen H.H. 2008. Analysis of intention in dialogues using category trees and its application to advertisement recommendation. In *Proceedings of the Third International Joint Conference on Natural Language Processing (IJCNLP 2008)*, pp. 625-630, Hyderabad, India.
- Jimenez, S., Becerra C. and Gelbukh, A. 2013. UNAL: Discriminating between Literal and Figurative Phrasal Usage Using Distributional Statistics and POS Tags. In *Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pp. 114-117.
- Jung, Y., Choi, Y. and Myaeng, S.H. 2007. Determining Mood for a Blog by Combining Multiple Sources of Evidence. In *Proceedings of International*

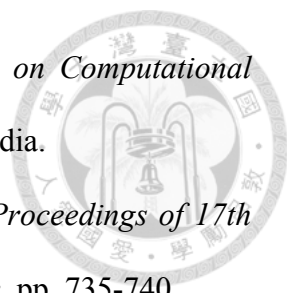
- 
- Conference on Web Intelligence*, pp. 271-274.
- Katz, G. and Giesbrecht, E. 2006. Automatic Identification of Non-compositional Multi-word Expressions Using Latent Semantic Analysis. In *Proceedings of the ACL/COLING-06 Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*, pp. 12-19, Sydney, Australia.
- Kaufmann, M. 2010. Syntactic Normalization of Twitter Messages. In *International Conference on Natural Language Processing*. Kharagpur, India
- Kilgarriff, A. and Rose, T. 1998. Measures for Corpus Similarity and Homogeneity. In *Proceedings of 3rd Conference on Empirical Methods in Natural Language Processing*, pp. 46-52. Granada, Spain.
- Ku, L.W. and Chen, H.H. 2007. Mining Opinions from the Web: Beyond Relevance Retrieval. *Journal of American Society for Information Science and Technology*, 58(12): 1838-1850.
- Lafferty, J., McCallum, A. and Pereira F. 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the Eighteenth International Conference on Machine Learning*, pp. 282–289.
- Lee, Chang-Ye. 2011. Study of recommending friends to organizations in microblog platform. Master thesis. National Taiwan University.
- Li, L. and Sporleder, C. 2010. Linguistic Cues for Distinguishing Literal and Non-Literal Usages. In *Proceedings of 23rd International Conference on Computational Linguistics (COLING 2010), Poster Volume*, pp. 683-691, Beijing, China.
- Li, Z. and Yarowsky D. 2008. Mining and Modeling Relations between Formal and Informal Chinese Phrases from Web Corpora. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

- 
- Liebrecht, C., Kunneman, F. and Bosch, A. 2013. The Perfect Solution for Detecting Sarcasm in Tweets #not. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pp. 29-37, Atlanta, Georgia.
- Lin, H.Y., Yang, C.H. and Chen, H.H. 2007. What Emotions Do News Articles Trigger in Their Readers? In *Proceedings of 30th Annual International ACM SIGIR Conference*, pp. 733-734.
- Lin, H.Y. and Chen, H.H. 2008. Ranking Reader Emotions Using Pairwise Loss Minimization and Emotional Distribution Regression. In *Proceedings of 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 136-144.
- Lin H.Y., Yang, C.H. and Chen, H.H. 2008. Emotion Classification of Online News Articles from the Reader's Perspective. In *Proceedings of International Conference on Web Intelligence*, pp. 220-226.
- Liu, B. 2012. Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, Morgan & Claypool Publishers.
- Liu, Y., Huang, X., An, A. and Yu, X. 2007. ARSA: A Sentiment-Aware Model for Predicting Sales Performance Using Blogs. In *Proceedings of the 30th Annual International ACM SIGIR Conference*, pp. 607-614.
- Lukin, S. and Walker, M. 2013. Really? Well. Apparently Bootstrapping Improves the Performance of Sarcasm and Nastiness Classifiers for Online Dialogue. In *Proceedings of the Workshop on Language Analysis in Social Media*, pp. 30-40, Atlanta, Georgia.
- Mei, J., Zhu, Y., Gao, Y. and Yin, H. 1982. *Tóngyìcílín*. (同義詞詞林) Shanghai Dictionary Press.
- Mishne, G. 2005. Experiments with Mood Classification in Blog Posts. In *Proceedings*

- of 1st Workshop on Stylistic Analysis of Text for Information Access.*
- Mishne, G. and Rijke, M. De. 2006. Capturing Global Mood Levels Using Blog Posts. In *Proceedings of AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs*, pp. 145-152.
- Mukherjee, A., Liu, B. and Glance N. 2012. Spotting Fake Reviewer Groups in Consumer Reviews. In *Proceedings of the International World Wide Web Conference (WWW 2012)*, pp. 191- 200.
- Mukherjee A., Kumar A., Liu, B., Wang, J., Hsu, M., Castellanos, M. and Ghosh, R. 2013. Spotting Opinion Spammers using Behavioral Footprints. In *Proceedings of SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2013)*, pp. 632-640.
- Ott M., Choi, Y., Cardie, C. and Hancock, J. 2011. Finding Deceptive Opinion Spam by Any Stretch of the Imagination. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pp. 309–319.
- Pang, B., Lee, L. and Vaithyanathan, S. 2002. Thumbs up? Sentiment Classification Using Machine Learning Techniques. In *Proceedings of 2002 Conference on Empirical Methods in Natural Language Processing*, pp. 79-86.
- Rayson, P. and Garside R. 2000. Comparing Corpora Using Frequency Profiling. In *Proceedings of Workshop on Comparing Corpora of ACL 2000*, pp. 1-6.
- Reyes, A., Rosso, P. and Buscaldi, D. 2012. From Humor Recognition to Irony Detection: *The Figurative Language of Social Media. Data & Knowledge Engineering*, 74:1-12.
- Sahami, M. and Heilman, T.D. 2006. A Web-based Kernel Function for Measuring the Similarity of Short Text Snippets. *WWW'06*, pp. 377-386. ACM Press.
- Scaiano, M. and Inkpen, D. 2011. Finding Negative Key Phrases for Internet

- 
- Advertising Campaigns Using Wikipedia. In *Recent Advances in Natural Language Processing (RANLP 2011)*, pp. 648–653, Hissar, Bulgaria.
- Sperber, D. and Wilson, D. 1981. Irony and the Use-Mention Distinction. In *Radical Pragmatics*, pp. 295-318.
- Sperber, D. and Wilson, D. 1992. On Verbal Irony. In *Lingua* 87, pp. 53-76.
- Sporleder, C. and Li, L. 2009. Unsupervised Recognition of Literal and Non-Literal Use of Idiomatic Expressions. In *Proceedings of the 12<sup>th</sup> Conference of the European Chapter of the ACL*, pp. 754-762.
- Sun Y.T., Chen, C.L., Liu C.C., Liu, C.L. and Soo V.W. 2010. Sentiment Classification of Short Chinese Sentences. In *Proceedings of 22nd Conference on Computational Linguistics and Speech Processing*, Nantou, Taiwan, pp. 184-198.
- Strapparava, C. and Mihalcea, R. 2007. Affective Text. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pp. 70–74.
- Tang, Y.J, Chang, Y.L. and Chen, H.H. 2011. A Comparison between Microblog Corpus and Balanced Corpus from Linguistic and Sentimental Perspectives. In *Analyzing Microtext: Papers from the 2011 AAI Workshop*.
- Utsumi, A. 1996. A Unified Theory of Irony and Its Computational Formalization. In *Proceedings of the 16th Conference on Computational Linguistics*, pp. 962-967.
- Vanzo A., Croce D. and Basili R. 2014. A Context-based Model for Sentiment Analysis in Twitter. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. Pp. 2345-2354. Dublin, Ireland.
- Veale, T. and Hao, Y.F. 2010. Detecting Ironic Intent in Creative Comparisons. In *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI 2010)*, pp. 765-770, Lisbon, Portugal.
- Wang, F., Wu, Y.F. and Qiu, L.K. 2012. Exploiting Discourse Relations for Sentiment



- 
- Analysis. In *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012): Posters*, pp. 1311-1320, Mumbai, India.
- Wiebe, J. 2000. Learning Subjective Adjectives from Corpora. In *Proceedings of 17th Conference of the American Association for Artificial Intelligence*, pp. 735-740.
- Xia, Y.Q., Wong, K.F. and Gao W. 2005. NIL is not Nothing: Recognition of Chinese Network Informal Language Expressions. In *Proceedings of the 4th SIGHAN Workshop on Chinese Language Processing*, pp. 95-102.
- Yang C.H., Lin, H.Y. and Chen H.H. 2007a. Building Emotion Lexicon from Weblog Corpora. In *Proceedings of 45th Annual Meeting of Association for Computational Linguistics*, pp. 133-136.
- Yang, C.H., Lin, H.Y. and Chen, H.H. 2007b. Emotion Classification Using Web Blog Corpora. In *Proceedings of 2007 IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 275-278.
- Yang, C.H., Lin, H.Y. and Chen, H.H. 2008. Sentiment Analysis in Weblog Using Contextual Information: A Machine Learning Approach. *International Journal of Computer Processing of Languages*, 21(4): 331–345.
- Yang, C.H., Lin, H.Y. and Chen, H.H. 2009. Writer Meets Reader: Emotion Analysis of Social Media from both the Writer's and Reader's Perspectives. In *Proceedings of International Conference on Web Intelligence*, pp. 287-290.
- Yeh, Ming-kung. 2014. Weekly Food and Drug Safety. No. 440. February, Food and Drug Administration, Taiwan. Available at <http://www.fda.gov.tw/TC/PublishOther.aspx>
- Zhou, L., Li, B.Y., Gao, E., Wei, Z.Y. and Wong, K.F. 2011. Unsupervised Discovery of Discourse Relations for Eliminating Intro-Sentence Polarity Ambiguities. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language*

*Processing (EMNLP 2011)*, pp. 162-171, Edinburgh, Scotland, UK.

