

# Classifying mood in plurks

Mei-Yu Chen\*  
meifish.kat@gmail.com

Hsin-Ni Lin\*  
hnl98@std.ntnu.edu.tw

Chang-An Shih\*\*  
cashih0805@gmail.com

Yen-Ching Hsu\*  
depend2030@gmail.com

Pei-Yu Hsu\*  
moon80206@gmail.com

Shu-Kai Hsieh\*  
shukai@gmail.com

\*Department of English, National Taiwan Normal University

\*\* Institute of Linguistics, Academia Sinica

## Abstract

In this paper, we present a simple but efficient approach for the automatic mood classification of microblogging messages from Plurk platform. In contrast with Twitter, Plurk has become the most popular microblogging service in Taiwan and other countries<sup>1</sup>; however, no previous research has been done for the emotion and mood recognition, nor the Chinese affective terms or corpus available. Following the line of mashup programming, we thus construct a dynamic plurk corpus by pipelining Plurk APIs, Yahoo! Chinese segmentation APIs, etc to preprocess and annotate the corpus data. Based on the corpus, we conduct experiments by way of combining textual statistics and emoticons data, and our method yield the results with high performance. This work can be further extended to combine with affective ontology designed with emotion theory of appraisal.

Keyword: mood classification, plurks, keyness, emotion paradox

---

<sup>1</sup> According to Alvin, the cofounder of Plurk website, the number of the plurkers in Taiwan had reached approximately 1 million, which was one-third of the total plurkers in October, 2009. Another statistic data is collected from *Google trend for website*, manifesting that Taiwan is the rank one region of visiting Plurk website (August,2010).

## 1. Introduction

Recently, with the emergence of Web 2.0 content, sentiment and opinion mining in social media and network domains like Blogs, has gained great attention from computational linguistics and other fields. Mood/Emotion classification, as related to sentiment classification, whose goal is set to tag a text based on whether it expresses positive or negative sentiment, has been surveyed on lyrics in music domain.

Mood classification is useful for various applications; the mood intelligence module can be incorporated in (affective) dialogue systems, learning tutors, etc. to improve the naturalness of human-machine interaction.

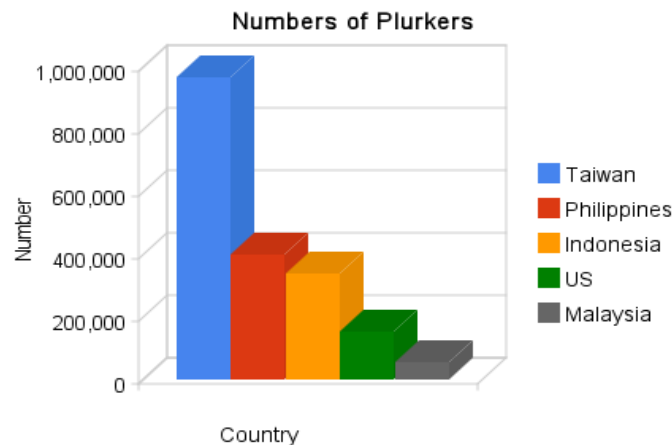


Figure 1. Number of Plurkers

This paper addresses the task of classifying microblogging posts by mood. What motivates us to conduct this research is the fact that the above figure shows: Plurk<sup>2</sup> is the most commonly used microblogging platform in Asia, and in contrast with other social networks, it has also become one of the most common forms of computer-mediated communication in Taiwan. In this paper, we aim to propose a simple and effective method in automatic classifying emotion of plurk messages for emotion.

The rest of the paper is organized as follows. Section 2 introduces the *emoticons* in plurk messages. Section 3 describes the construction of plurk corpus. Section 4 presents the experiment results and evaluation. Finally, Section 5 concludes the paper and summarizes our future work.

## 2. Emoticon and Emotional Expression in Social Media

In the previous studies, computer systems have been designed to perform automatic sentiment analysis using elements from machine learning such as latent semantic analysis (LSA), Naive Bayes, support vector machines (SVM) etc. [1] [2]. A different set of features was utilized. For example, Mishne [1] uses many features extracted from Live Journal web blog service to train the SVM binary classifier for sentiment analysis.

---

<sup>2</sup> <http://www.plurk.com>

It becomes even harder in the task of identifying (multiple) emotions. Jung et al. [3] show that there are some idiosyncratic natures of mood expression in plurk messages; for example, the initial mood may not be maintained all the way to the end (which also known as the fluctuation of moods). In addition, some blogs are so intertwined that even human readers would have difficulties in indentifying the mood, not to mention for the machine.

By scrutinizing the plurk data, we found that the *emoticons*<sup>3</sup> can be utilized as useful features in presenting the emotional content of a conversation, especially in computer-mediated communication (CMC). As graphic representation of facial expressions, emoticons are used to remedy the lack of some non-verbal cues, such as head nodding, smiles, and it is why they are often imbued in the textual messages, where we want to set different tones, such as humor, irony, sarcasm, cuteness, flippancy or other emotions that our face is not there to deliver [4].

With the advantages of that fact that plurk messages are augmented with many plugins support for emoticons, we propose to extract the affective terms by 'bag-of-words' bootstrapped from the emoticons (non-verbal expressions) widely used in plurk domains, and use them to determine the overall emotion embedded within a plurk message.

### 3. Plurking a Social Web Corpus

For this research, we constructed a Plurk social web-as-corpus (SocialWaC) in Taiwan. Plurk is a social journal service that allows users to showcase the events that make up their life, and follow the events of the people that matter to them in messages called *plurks*. Plurkers just post a new plurk and it'll appear on the timeline.

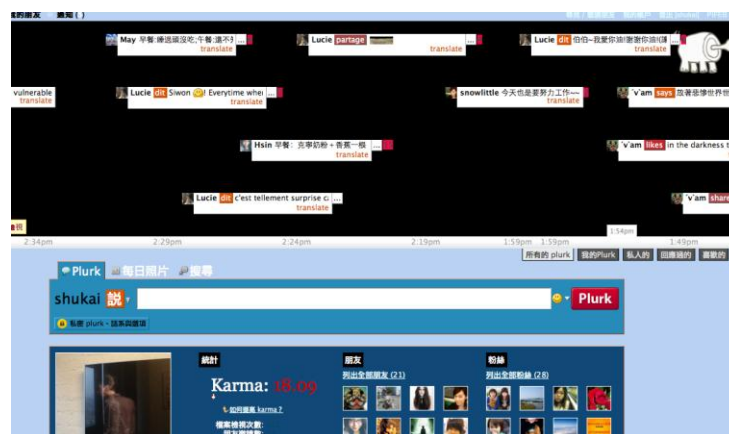


Figure 2. Snapshot of plurk timeline

### 3.1 Crawling and Preprocessing

#### 3.1.1 Crawling Data

We use the official Plurk API<sup>4</sup> to crawl the Plurk data. These include *plurk data* (with attributes of plurk\_id, qualifier, date when the plurk was posted, content, response, emoticons,

<sup>3</sup> Note that in this paper, we do not consider *smileys*, which is often used in other CMCs like emails communication.

<sup>4</sup> <http://www.plurk.com/API>

etc), *user data* (with attributes of id, location, date\_of\_birth, gender, karma, relationship, etc), and other *social network data* (friends and fans, blocks, cliques, etc).

We stick to the policy of Plurk with regards to which technical, conversational and sociological *meta-information* should be captured during the collection of the data and which ethical aspects should be considered when designing our plurk SocialWaC. For instance, privacy data were not collected. Currently, our corpus has 3000 plurkers, with 1723 female plurkers and 1277 male plurkers. The total number of plurks achieved to 38629.

### 3.1.2 Format

The API returns JSON encoded data. We use simplejson of Python library to decode the data returned. We decided to crawl the data without layout information (such as typeface, font, color, size, etc) for some technical reasons on the one hand, and on the other hand, we found that plurkers perform pragmatic functions more often than just text decoration within communicative exchanges, except the emoticons, whose visual properties are functional in the original texts.

### 3.1.3 Segmentation

One of the most crucial and difficult tasks in preprocessing corpus in Chinese is the word segmentation. Among different segmentation algorithms, the *lexicon-based* approach is widely adopted, which can successfully identify Chinese sentences as distinct words from Chinese texts. However, the word identification ability of the lexicon-based scheme is highly dependent with a well-prepared lexicon with sufficient amount of lexical entries. Hybrid approach thus proposed to combine with other statistical information to detect out-of-vocabulary (OOV).

To our knowledge, most Chinese word segmentation systems, such as Sinica *CKIP segmentator* and *CScanner* were designed to perform Chinese word segmentation process for general domain Chinese texts based on a lexicon manually. For our social domain corpus with huge amount of lexical and syntactic varieties (new words, phrases and sentence structures), these tools did not work well. For instance, these tools cannot identify trendy words like *luo2 li4* (蘿莉) 'a little girl', *xing2 nan2* (型男) 'a stylish man', let alone those complex orthographically code-mixing words like *A ka* (A 咖) 'the most popular and famous guy in certain domains', *Q ban3* (Q 版) 'cute version', *niou2 B* (牛 B) 'great', etc. Based on these considerations, we adopt the segmentation system provided by Yahoo!<sup>5</sup> for the sake of its powerful lexicon extension for new words emerged in the social media.

## 3.2 Annotation

In contrast with traditional corpus annotation, King [5] pointed out that the responses to spam (in the form of 'adbots'), cyber-orthography, the ubiquity of names, and overlapping conversations all result in the challenges of annotation in CMC.

In addition to these properties in common, 'plurks' also behave as short sentences and

---

<sup>5</sup> <http://tw.developer.yahoo.com/cas/>

fragments, with abundant use of emoticon, and available meta-information of users. As emoticons convey the sentiment that plurkers use to express their particular state of mind, the annotation is based on plurk level, with POS information and emoticons tagged.

One particularly interesting property that can be observed from our corpus data, is the availability of users' meta-information, which can be coupled with lexical information for computational sociolinguistic surveys. For instance, Figure 3 shows a socialized lexical network extracted from our corpus, visualized by vister software. The colored areas demonstrate the shared lexical patterns among certain social community.



Figure 3. Socialized lexical network

In addition, through the meta-information, we are also able to know when a certain gender or people from a certain location use a lexicon, how accompanying emotion is distributed. This is demonstrated in the website introduced in Section 3.3.

### 3.3 Availability

We've made a website for the query of our corpus, which can be accessed via <http://140.122.83.235/plurk/>.

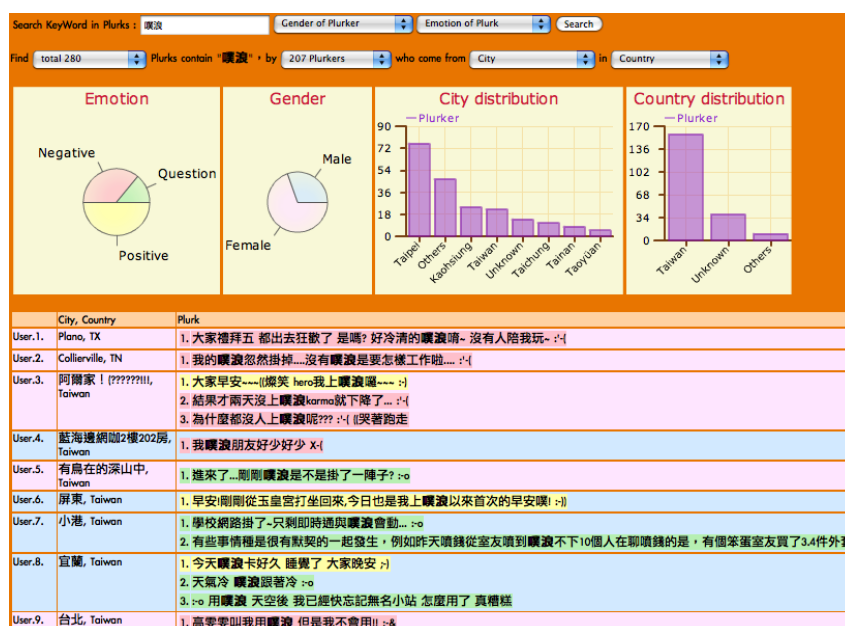


Figure 4. Plurk search website

## 4. Experiments

In this section we describe the mood classification experiments we conducted on plurk corpus. The aim was to assess affective and attitude interpretations of plurks based on our proposed method.

### 4.1 Algorithm: Keyness and Manual Validation

The prevalent SVM machine learning algorithm works well for text classification, but has only low precision and recall for sentence-level classification. Various techniques need to be used to deal with the issue of imbalanced data distribution [6].

In this paper, we use a simple but effective method that ties back to a well-known textual statistical measure called *keyness*. Given a set of evaluative plurks  $\mathcal{P}$ , our mood classifier classifies each plurk  $p \in \mathcal{P}$  one of the three affective valencies: *positive*, *interrogative* and *negative*. This algorithm makes use of Yahoo! Chinese segmentator and part-of-speech (POS) tagger, with a stop-word list compiled from Academia Sinica Balanced Corpus. It consists of four steps explained in the following:

**[Step one]** Extract plurks containing emoticons from the corpus and divide them into 3 different emoticon categories based on the emoticons they contain, so as to generate the keyword and keyness lists as in step 2.

**[Step two]** Generate a list of keyword and keyness for each emoticon categories respectively, i.e. Positive (P), Interrogative (I), and Negative (N) categories,<sup>6</sup> by using the *log-likelihood* feature selection method in *Anctconc* [7][8].<sup>7</sup> In addition, stop words<sup>8</sup> are removed from the keyword lists.

We define a **Saliency Score (SS) Function**, which is calculated by the result of one keyword's keyness value divided by the keyness value of the first-ranking word in the keyword list, as shown in (1).

$$SS = \frac{Keyness(KW)}{Keyness(Rank\_1KW\ of\ Category)} \quad (1)$$

**[Step three]** Calculate the sum of saliency score of each emoticon category in the whole sentence domain, the category with the highest sum is assigned, i.e., *Emoticon assigned* ( $X$ ) =  $Max (Score(P), Score(I), Score(N))$ . Unknown words are assigned to the most similar words using **pointwise mutual information (PMI)** from Chinese Wordnet.

$$PMI(w_1, w_2) = \log_2 \frac{P(w_1 \wedge w_2)}{P(w_1)P(w_2)} \quad (2)$$

---

<sup>6</sup> Note here that emoticons themselves are not counted as a word when we try to generated keywords. Emoticons are used so that we can categorized plurks as positive, interrogative, or negative.

<sup>7</sup> When generating the list of keyword and keyness of Positive category, we take the other categories as reference corpus. The situation is similar when we generate the lists of Interrogative and Negative categories respectively.

<sup>8</sup> Compiled from Academia Sinica Balanced Corpus. The Stop lists contain the deictics, pronouns and quantity words.

**[Step four] Manual Validation:** After the automatic emoticon assignment, a manual validation is conducted. Two evaluators are requested to validate the emoticon assignment. This step aims to not only validate the results but also recall the accurate judgment of the aforementioned automatic emotion classification.

## 4.2 Data and Resource Description

To examine emoticons and emotional expression, we extract plurks in our corpus. Of 436,487 plurks, 103,171 (24%) were found that contain emoticons, 32,851 found containing default emoticons. To simplify the classification of emoticon categories, only default emoticons<sup>9</sup> are concerned in this current research.

We randomly divide the data with default emoticons into 80/20 division for training and testing, respectively. The keyword and keyness lists of 3 different emoticon categories are generated by the 80 division, and the testing data are construed by the 20 division.

## 4.3 Experiments with heuristic rules (H-rules)

All the test plurks are assigned to one of the three valences by the method described in Section 4.1. The second and third experiment is conducted with the following heuristic rules, respectively.

### 4.3.1 H1-Rule: Decomposing the plurks

The focus of our proposed method is similar to sentiment classification at the *sentence-level*. However, according to the *Pragmatic Information Structure*, new information, the purpose of communication, is conveyed later than old information in sentences. As Emoticon possessing the communicative function, its regular sentence final position is corresponded to the *Information Structure*, and is closely related to the later fragment of sentence, especially in disjunctive sentence. For example, “Fragment 1 [I won the lottery this morning,] Fragment 2 [ but I lost the ticket in the afternoon ☹.]” The fragment 1 conveys a proposition of positive emotion, but the real conveying information of this sentence is the second fragment, conveying a negative emotion and denoting by the emoticon in sentence final position. We thus wish to examine whether the sentence final fragment is most probable to contain affective content by restrict the Score Domain to the sentence final fragment only.

### 4.3.2 H2-Rule: Deleting inappropriate keywords

A few words cannot serve as Negative keywords due to the properties of their meanings; therefore, we delete those content words in our keyword lists. There are 3 words being deleted.

The first one is *hao3* (好) 'good (adj.)/very (adv.)'. As shown in the English translation, *hao3* (好) can be a word of positive emotion ( i.e. "good") or an adverb of degree

---

<sup>9</sup> Default emoticon can be accessed by all users of different Karma value. In Plurk, the Karma value is a mechanism used to evaluate user's level.

(i.e. "very"), which conveys neutral emotion but usually be used in negative emotion to amplify the negative degree.

The second keyword being deleted is *le* (了), a Chinese aspectual particle. The particle provides no meaning but aspectual information. Therefore, *le* (了), like stop words, was not supposed to be a keyword.

The other word being deleted is *chu fu2* (舒服) 'comfortable', a word is positive per se but is misanalyzed as a keyword of negation emotion. The misanalysis appears in several positive words in the experiment because those words usually follows *bu4* (不) 'not', a negation adverb, to form a negative meaning. A problem arises here is that Yahoo! Chinese Segmentation system segmented the *bu4*-plus-word(s) pattern inconsistently. Sometimes *bu4* is separated with the following adjective or verb; sometimes the *bu4* and the adjective or verb together are segmented as a single word. To avoid the problem caused by inconsistency, we examine the 30-most-frequent separated words following *bu4* in negative keywords. Among the 30 keywords, only *chu fu2* 'comfortable' is a word with positive emotion, so we delete only the word here.

#### 4.4 Baseline

Given no similar experiments have been conducted before, the baseline of our experiments is set as the results of randomly classifying each plurk in the testing data as one of valences (i.e. positive, interrogative, or negative). The random classification is automatically performed by a program written in Python.

#### 4.5 Results and Analysis

The results of the three experiments described in Section 4.1 and 4.3 are illustrated in Table1.

Table 1. The results of 3 experiments with and without heuristic rules

	Experiment 1 (without H-rules)	Experiment 2 (H1-rule)	Experiment 3 (H2-rule)
Accuracy	59.13 %	56.35 %	61.80 %
Recall	95.05 %	91.15 %	94.55 %

From the numbers of Table 1, it is observed that Experiment 3 has the best performance among all. For a more convincing result, we duplicate the Experiment 3 to examine whether the heuristic rule of Keywords is the essential factor which affects the result. The duplicated experiment (Experiment 4) is conducted by randomly choosing another 20 division of testing data. The result is illustrated in Table 2, and recalls by *manual validation (MV)* and *PMI* are added in the table.



Table 2. The results of 2 experiment both with H2-rule

	Experiment 3 (H2-rule)		Experiment 4 (H2-rule)	
	Original	MV +PMI	Original	MV+PMI
Accuracy	61.80 %	70.89 %	60.64 %	69.55 %
Recall	94.55%	95.3%	94.2%	94.98%
Ambiguous <sup>10</sup>		11.02%		10.90 %

The detailed accuracy rate of distinct emoticon categories of Experiment 3 and 4 is illustrated in Table 3 and Table4.

Table 3. The accuracy rate of distinct emoticon categories in Experiment 3

	Experiment 3 (H2-rule)					
	Positive (n= 3122)		Interrogative (n= 460)		Negative (n= 2779)	
	Original	MV +PMI	Original	MV +PMI	Original	MV +PMI
Accuracy	56.95%	59.39%	25.65%	60.22%	73.22%	85.50%
Recall	93.69%	94.59%	98.04%	98.7%	94.93%	95.54%

Table 4. The accuracy rate of distinct emoticon categories in Experiment 4

	Experiment 4 (H2-rule)					
	Positive (n= 3563)		Interrogative (n= 516)		Negative (n= 3093)	
	Original	MV +PMI	Original	MV +PMI	Original	MV +PMI
Accuracy	56.55%	59.05%	19.38%	52.71%	72.23%	84.45%
Recall	93.49%	94.5%	95.16%	95.54%	94.86%	95.44%

Besides, the testing data of Experiment 3 and Experiment 4 are also used as the data of the baseline in this study; also, in order to produce a convincing results, we perform the baseline experiments twice. The results are presented in the following table.

<sup>10</sup> The inconsistency evaluation of two evaluators in manual validation, that is, while some sentences have ambiguous emotions, the evaluators have different agreements in the emoticon assignment. The consistency rate of the two evaluators is 67.26 %.

Table 5. The results of baseline experiments

	Baseline of Experiment 3		Baseline of Experiment 4	
	1 <sup>st</sup> time	2 <sup>nd</sup> time	1 <sup>st</sup> time	2 <sup>nd</sup> time
Accuracy	34.73 %	32.86 %	32.54 %	33.73 %
Failure	65.27 %	67.14 %	67.46 %	66.27 %

Compared the accuracy in Table 5 with that in Table 2, it is shown that the results of our algorithm are much higher than those of baseline, especially the recalled results.

#### 4.6 Evaluation

From the result in Table 1, heuristic rule 1, decomposition the plurk, did not perform better as we expected. Instead, the accuracy rate decreases from 59.13% to 56.35% and recall rate decreases from 95.05% to 91.15%. This result can be contributed to the shorter length of fragments, not enough keywords to be extracted from. However, heuristic rule 2, Deleting inappropriate keywords, increases the accuracy rate from 59.13% to 61.80%, and did not affect much of the recall rate. We proposed that some special categories of keywords should be removed. First is the homographic keywords, which can ambiguously serve as keywords in two polar emotions, such as *hao3* (好). Second is the positive adjective which can combine with the negative adverb to denote an opposite meaning. As the homographic keywords, these positive terms can serve as keywords in two polar emotions, as positive terms or negative when appearing after negative terms. A plausible solution is to retag the negative expression, combining the negative term and the following positive term into a single word, to avoid undetermined **semantic orientation** [9].

It is worth noting that researchers found that "in relationship with simultaneous verbal behavior, nonverbal behavior may emphasize, repeat, substitute, or contradict verbal messages, yet CMC commentators discuss emoticons in terms of their emphatic function or signaling function, not mere repetition or substitution of otherwise-conveyable verbally transmitted meaning". So basically, "positively valenced emoticons should enhance positively valenced verbal messages, and negative emoticons make negatively valenced messages more negative". However, there exists an interesting phenomenon of **emoticon paradox** which suggests an intentionally conflicted or ambiguous state, and which are thus less predictable. That is, "positive verbal messages with a negative emotion or vice versa". For example, "*I really enjoy the meal with my boss 😞.*" Verbally, the speaker said that he enjoys, but he conveys his real feeling by using negative emoticon. The inconsistencies between verbal meanings and nonverbal cues conveyed by emoticons will render an inconsistency error in classification task. This phenomenon can be observed in the accuracy of Interrogative and Negative categories in Table 3 and Table 4; after manual validation, the accuracy of the two categories improves greatly due to the hedge-of-speech-act phenomenon. We observe that when conveying a negative expression, people like to hedge the speech act by adopting

interrogative format. For example, “*How come there are such rude people in the world? ☹*” On the other hand, negative expression could be hedged or modified by using an interrogative emoticon, such as the interrogative emoticon. For example, “*I don’t want to take the exam. :-o.*” However, it is rarely to see a hedge in the positive expression, thus the accuracy rate didn’t improve much after manual validation.

We propose that the emoticons not only serve as a decoration or amplification of the sentence mood, they also function as verbal tokens conveying their arbitrary meanings as observed in **emotion paradox**. Even without any content, they convey information by themselves. The emotion paradox phenomenon can be used to automatically detect the pragmatic hedge-of-speech-act usage. When analyzing sentence like “*I really enjoy the meal with my boss ☺.*” By detecting the inconsistency of the computed emoticon and the given emoticon, the hedge-of-speech-act usages can be detected and contribute to the real meaning of the sentence, which cannot be done by purely analyzing plain verbal meanings.

## 5. Conclusion

The present study proposed a simple but effective way, by computing the keyness of keywords in sentence, in both assigning emotions to sentences and detecting the emotion paradox sentences. This method can as well compile a dynamic (Chinese) affective terms, the keywords, and apply to subfields of Cognitive linguistic, Chinese Teaching and so forth. Pragmatics will also strongly benefit from the detection of the hedge-of-speech-act sentences. Since in the traditional methodology, the presuppositions or implicatures in utterances are known by speakers only. However, an emoticon provides a cue to speakers’ real emotion and intention, and thus reflects speakers’ underlying meaning of presuppositions and implicatures. Future works can be aimed at using LSA to broaden the linkage of the keywords to see whether the closely related words sharing the same emotion, and whether this approach can expand the coverage of keywords to reduce achieving a lower unknown rate.

## References

- [1] G. Mishne, “Experiments with Mood Classification in Blog Posts,” in *Proceedings of the 1st Workshop on Stylistic Analysis of Text For Information Access*, 2005.
- [2] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” Dec 2009. [Online]. Available: <http://www.stanford.edu/~alecmgo/papers/TwitterDistantSupervision09.pdf>
- [3] Y. Jung, Y. Choi, and S.H. Myaeng, “Determining mood for a blog by combining multiple sources of evidence,” in *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 271-274, 2007.
- [4] J. B. Walther, and K.P. D'addario, “The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication,” *Social Science Review*, vol. 19, no. 3, pp. 324-347, 2001.
- [5] B. King, “Building and Analysing Corpora of Computer-Mediated Communication,” in *Contemporary Corpus Linguistics*, P. Baker, Ed. New York: Continuum International Publishing Group, 2009, pp. 301-320.

- [6] E. Spyropoulou, S. Buchholz, and S. Teufel, "Sentence-based Emotion classification for text-to-speech synthesis," presented at Computational Aspects of Affectual and Emotional Interaction-2008, Patras, Greece, 2008.
- [7] L. Anthony, "AntConc: A Learner and Classroom Friendly, Multi-Platform Corpus Analysis Toolkit," in *Proceedings of IWLeL 2004: An Interactive Workshop on Language e-Learning*, pp. 7-13, 2004.
- [8] A. Kilgarriff, "Comparing corpora," *International Journal of Corpus Linguistics*, vol. 6, no. 1, pp. 97-133, 2001.
- [9] P. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation applied to Unsupervised Classification of Reviews," in *Proceeding of the Meeting of the Association for Computational Linguistics*, pp. 417-424, 2002. [Online]. Available: <http://acl.ldc.upenn.edu/P/P02/P02-1053.pdf>