

# NTOUA at IJCNLP-2017 Task 2: Predicting Sentiment Scores of Chinese Words and Phrases

Chuan-Jie Lin and Hao-Tsung Chang

Department of Computer Science and Engineering  
National Taiwan Ocean University  
Keelung, Taiwan ROC.

{cjlin, htchang.cse}@ntou.edu.tw

## Abstract

This paper describes the approaches of sentimental score prediction in the NTOU DSA system participating in DSAP this year. The modules to predict scores for words are adapted from our system last year. The approach to predict scores for phrases is keyword-based machine learning method. The performance of our system is good in predicting scores of phrases.

## 1 Introduction

The task of Dimensional Sentiment Analysis for Chinese Phrases (DSAP), a shared task held in IJCNLP 2017, focuses on predicting valence and arousal scores of Chinese words and phrases. It is the second evaluation project of sentiment score prediction in word level (Yu *et al.*, 2016b), but the first task in phrase level.

The valence of a word or phrase represents the degree of pleasant and unpleasant (or positive and negative) feelings. For example, “happy” is a positive word and “sad” is negative.

The arousal of a word or phrase represents the degree of excitement and calm. For example, “surprise” is more excited and “tired” is calmer.

Valence and arousal can be used to define a space where the features denote dimensions (Russell 1980; Kim *et al.*, 2010; Malandrakis *et al.*, 2011; Wei *et al.*, 2011; Calvo and Kim, 2013; Paltoglou *et al.*, 2013; Yu *et al.*, 2015; Wang *et al.*, 2016). These two dimensions are independent. There are positive-excited words like “delighted”, positive-calm words like “relaxed”, negative-excited words like “angry”, and negative-calm words like “bored”.

The applications of sentiment analysis include antisocial behavior detection (Munezero *et al.*, 2011), mood analysis (Choudhury *et al.*, 2012) and product review ranking (Ren and Nickerson, 2014).

It was our second attempt of sentiment rating prediction in word level. We simply chose the best two systems developed during DSAW in 2016 (Yu *et al.*, 2016a),

which are described in Section 2. For phrases, we proposed a simple keyword-based machine learning method as described in Section 3.

## 2 Predicting by Co-Occurrence

This year, we experimented two simple methods to predict sentiment scores of words. One method focuses on predicting valence scores and the other on arousal scores. Both systems use co-occurrence information from Google Web 1T 5-grams<sup>1</sup> (Google N-grams for short hereafter).

To illustrate our method, we first define two functions of co-occurrence scores between a target word  $t$  and a context word  $w$ . The *right co-occurrence frequency*  $coFreqR(t, w, n)$  is the frequency of the  $n$ -gram  $a_1 a_2 \dots a_n$  where  $a_1 = t$  and  $a_n = w$ . The *left co-occurrence frequency*  $coFreqL(t, w, n)$  is the frequency of the  $n$ -gram  $a_1 a_2 \dots a_n$  where  $a_1 = w$  and  $a_n = t$ . Note that we only used bigram to 5-gram data, therefore  $2 \leq n \leq 5$ .

### 2.1 Co-Occurrence with Sentiment Words

The first system predicts sentiment scores of a target word by its co-occurrence with other sentiment words. All sentiment words provided in the DSAW training data are considered as the “context words” in the  $coFreq()$  functions. Two kinds of features are defined as follows.

The *right co-occurrence sentiment features*  $sf_{nR}$  of a target word  $t$  is the average of

$$sentiScore(w) \times \log(coFreqR(t, w, n))$$

for all the sentiment words  $w$  whose  $coFreqR(t, w, n)$  values are positive, where  $sentiScore(w)$  is either the valence or arousal score of  $w$ .

Similarly, the *left co-occurrence sentiment features*  $sf_{nL}$  of a target word  $t$  is the average of  $sentiScore(w) \times \log(coFreqL(t, w, n))$  for all the sentiment words  $w$  whose  $coFreqL(t, w, n)$  values are positive. Given that  $2 \leq n \leq 5$ , totally four right co-occurrence sentiment features and four left co-occurrence sentiment features are defined.

<sup>1</sup> <https://catalog.ldc.upenn.edu/LDC2006T13>

## 2.2 Co-Occurrence with Degree Adverbs

The second system predicts sentiment scores of a target word by its co-occurrence with degree adverbs such as “非常” (very) and “有點” (a little). We collected a set of degree adverbs from Tongyici Cilin, a dictionary about Chinese synonyms. These degree adverbs are considered as the “context words” in the `coFreq()` functions. Two kinds of features are defined as follows.

The *right co-occurrence degree features*  $df_{nR}$  of a target word  $t$  is the average of  $\log(\text{coFreqR}(t, d, n))$  for all the degree adverbs  $d$  whose  $\text{coFreqR}(t, d, n)$  values are positive. Similarly, the *left co-occurrence degree features*  $df_{nL}$  of a target word  $t$  is the average of  $\log(\text{coFreqL}(t, d, n))$  for all the degree adverbs  $d$  whose  $\text{coFreqL}(t, d, n)$  values are positive. Given that  $2 \leq n \leq 5$ , totally four right co-occurrence degree features and four left co-occurrence degree features are defined.

## 3 Sentiment Changing by Adverbs

This is the first evaluation project of sentiment score prediction in phrase level. The “phrases” used this year have the same pattern:  $[\text{RB} \mid \text{MD}]^+ \text{JJ}$ , where JJ is an adjective, preceded by one or more adverbs (RB) or modals (MD). It is interesting to see how an adverb or modal word can change the sentiment score of an adjective.

There are two types of adverbs seen in the test data this year: negation words (NEG) and adverbs of degree (DEG). A negation word such as “不” (not) often alters a sentiment score into its opposite direction in some degree, while an adverb of degree such as “非常” (very) often enhance a sentiment score along its original direction.

However, the effect of modal words (MOD) such as “應該” (should) or “可能” (might) is less predictable. Maybe they would change the sentiment degrees toward the neutral point.

These adverbs and modal words can be compound as well. There are 4 combinations seen in the training data:

- DEG\_NEG: such as “完全 不 怕” (totally not afraid)
- MOD\_DEG: such as “可能 很 怕” (might-be very afraid)
- MOD\_NEG: such as “可能 不 怕” (might not be afraid)
- NEG\_DEG: such as “不 太 怕” (not very afraid)

Based on the adverbs or modal words detected before an adjective, we proposed two different methods to predict sentiment scores of phrases, as described in Sections 3.2 and 3.3. Before that, we will explain how we decide the preceding adverb combinations.

## 3.1 Adverb-Combination Detection

As we know that the phrases are written in the pattern of  $[\text{RB} \mid \text{MD}]^+ \text{JJ}$ , we use the following recursive grammars to detect the preceding adverbs and modal words.

```
sentiPhrase := [\text{RB} \mid \text{MD}] sentiPhrase  
sentiPhrase := [\text{RB} \mid \text{MD}] JJ
```

When parsing a phrase P, if it can be divided into an adverb (or modal) with an adjective (which appears in the dictionary), the division is accepted and the parsing is finished. Otherwise, if P can be divided into A+B where A is an adverb (or modal), B will be further parsed with the same grammars.

For example, the phrase “很難過” (very sad) is divided into “很” (very) and “難過” (sad), because “很” (very) is known as an adverb and “難過” (sad) is an adjective collected in the sentiment dictionary.

For another example, the phrase “沒有太難過” (not too sad) is divided into “沒有” (not) and “太難過” (too sad). Because “太難過” (too sad) is not a word, it is further divided into “太” (too) and “難過” (sad).

Note that the process can be repeated as many times as necessary with no limitation, so are our proposed methods. The longest combination found in the test data has only two words.

## 3.2 Phrase-Level Features

### 3.2.1 Deciding Core-Adjective Feature Values

The first two features we used are the valence and arousal scores of the core adjective. In the training process, only those training examples whose core adjectives can be also be found in the DSAW training set (which means their sentiment scores are correct) are used. In the testing process, if we do not know the sentiment scores of the core adjective, we will use our DSAW system to predict their scores in advance.

Note that both valence and arousal scores of the core adjective are used together, no matter when it is predicting the valence score or arousal score of a phrase.

### 3.2.2 Adverb-Combination Feature

The third feature we used is the adverb combination detected in front of the core adjective. We inserted underscores between adverbs, such as “應該\_沒有\_太”, “沒有\_太”, or “沒有” if only one adverb is found.

In another experiment, we only take the leftmost adverb as the feature value, which method is described in Section 3.3.2.

## 3.3 Phrase-Level Sentiment Prediction

We proposed two different systems to predict sentiment scores of phrases. The two systems worked on similar features but different procedures.

### 3.3.1 Predicting by Adverb-Combinations

The first system took the whole phrase of preceding adverbs and modal words as the third feature. Some examples of feature values are given here, where the first column depicts the phrases and the second the feature values:

更加小心 更加, 4.6, 6  
 稍微不小心 稍微\_不, 4.6, 6  
 稍微不痛 稍微\_不, 3, 6.2

In the first two examples, the core adjective is “小心” (careful) whose valence score is 4.6 and arousal score is 6, and the core adjective in the third example is “痛” (hurtful) whose valence score is 3 and arousal score is 6.2.

As mentioned in Section 3.2.1, if the sentiment scores of a core adjective is unknown (i.e. not found in the training data), its score will be predicted by DSAW system first in the testing process, or this example will be discarded during the training process.

### 3.3.2 Predicting by Single Adverbs

The second system only took the leading adverb or modal word as the third feature.

In the training process, these two sets of phrases were chosen as training data:

- Phrases in the pattern of [RB | MD] JJ whose core adjectives appear in the DSAW training set.
- Phrases in the pattern of [RB | MD] AA where AA is a phrase that appears in the DSAP training set.

In these two sets, the sentiment scores of the core adjectives or phrases are accurate thus can be used as feature values. There are two examples:

超級可愛 超級, 7, 6.2  
 超級不安全 超級, 3.444, 5

In the first example, the core adjective is “可愛” (cute), and the DSAW training data provides its valence score as 7 and arousal score as 6.2. In the second example, the core phrase is “不安全” (not safe), and the DSAP training data provides its valence score as 3.444 and arousal score as 5.

In the testing process, a given phrase is first divided into an adverb (or modal word) and a core phrase. If the sentiment scores of this core phrase can be found in the training data, the given phrase can be predicted directly.

If the sentiment scores of the core phrase are unknown, these scores should be predicted in advance, either by DSAP system or DSAW system. For example, to predict the phrase “超級不可恨”, it is divided into “超級” + “不” + “可恨” first. The sentiment scores of

the adjective “可恨” (hateful) is predicted by our DSAW system. And then the phrase “不” + “可恨” (not hateful) is predicted by our DSAP system with the just-predicted sentiment scores of “可恨” as feature values. Finally, the phrase “超級” + “不” + “可恨” (extremely not hateful) is predicted by our DSAP system by the sentiment scores of “不可恨” as feature values.

## 4 Run Submission and Evaluation Results

Two runs were submitted to the IJCLCLP 2017 Shared Task: Dimensional Sentiment Analysis for Chinese Phrases as requested by the organizers. Their strategies to train DSAW (for predicting words) and DSAP (for predicting phrases) modules are described as follows.

- NTOUA1: the DSAW module was trained by using co-occurrence sentiment features, and the DSAP was using trained by adverb-combination features.
- NTOUA2: the DSAW module was trained by using co-occurrence degree features; the DSAP was trained by using single-adverb features, and its prediction process was recursive as described in Section 3.3.2.

Both systems were trained by the random forest algorithm, a machine learning method. This method performed the best during our training process.

Table 1 lists the evaluation results of our submitted runs. Results were evaluated on (1) words (2) phrases (3) all about valence and arousal scores, respectively, in the metrics of mean absolute error (MAE) and Pearson correlation coefficient (PCC). The ranks of our systems among all the submitted systems are also depicted.

**Table 1. Performance of NTOU Runs in DSAP 2017**

Task & Metric (rank)	NTOUA1	NTOUA2
Word, Valence, MAE	0.913(15)	1.061(19)
Word, Valence, PCC	0.700(16)	0.544(22)
Word, Arousal, MAE	1.133(17)	1.114(16)
Word, Arousal, PCC	0.163(23)	0.350(21)
Phrase, Valence, MAE	0.472(7)	0.453(4)
Phrase, Valence, PCC	0.910(8)	0.929(5)
Phrase, Arousal, MAE	0.420(5)	0.441(6)
Phrase, Arousal, PCC	0.882(5)	0.870(6)
All, Valence, MAE	0.692(12)	0.757(13)
All, Valence, PCC	0.805(12)	0.737(15)
All, Arousal, MAE	0.777(11)	0.778(12)
All, Arousal, PCC	0.523(22)	0.610(12)
All, Rank	14.25	13

Both our DSAP modules perform very well in predicting sentiment scores of phrases (best ranked at Top 4).

There may not be significant difference between the two DSAP methods.

Simple adverb features are very effective. It also means that the sentiment score of the core adjective is the most critical information in sentiment prediction. This year, all the accurate sentiment scores of core adjectives can be found in the training data.

Unfortunately, neither of our DSAW modules achieved good performance in the formal test. We should greatly improve our DSAW module in the future.

## 5 Conclusion

This paper proposed two approaches to predict valence and arousal scores of Chinese words and two approaches for scores of Chinese phrases.

Before predicting a Chinese phrase, its leading adverbs or modal words are detected in advance. Its sentiment score can be predicted either by considering the whole set of leading adverbs, or recursively decided by single leading adverbs. The results show that this strategy achieves rather good performance.

Since the key to successfully predict sentiment scores of phrases is still a good DSAW system, we will study more about DSAW in the future.

## References

- Rafael Calvo, and Sunghwan Mac Kim. 2013. "Emotions in text: dimensional and categorical models," *Computational Intelligence*, 29(3): 527-543.
- Munmun De Choudhury, Scott Counts, and Michael Gamon. 2012. "Not all moods are created equal! Exploring human emotional states in social media," *Proceedings of ICWSM-12*, 66-73.
- Sunghwan Mac Kim, Alessandro Valitutti, and Rafael Calvo. 2010. "Evaluation of unsupervised emotion models to textual affect recognition," *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, 62-70.
- Su-Chu Lin, Shu-Ling Huang, You-Shan Chung, and Keh-Jiann Chen. 2013. "The lexical knowledge and semantic representation of E-HowNet," *Contemporary Linguistics*, 15(2): 177-194.
- Nikos Malandrakis, Alexandros Potamianos, Elias Iosif, and Shrikanth Narayanan. 2011. "Kernel models for affective lexicon creation," *Proceedings of INTERSPEECH-11*, 2977-2980.
- Myriam Munezero, Tuomo Kakkonen, and Calkin Montero. 2011. "Towards automatic detection of antisocial behavior from texts," *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP) at IJCNLP-11*, 20-27.
- Georgios Paltoglou, Mathias Theunis, Arvid Kappas, and Mike Thelwall. 2013. "Predicting emotional responses to long informal text," *IEEE Transaction Affective Computing*, 4(1): 106-115.
- Jie Ren and Jeffrey Nickerson. 2014. "Online review systems: How emotional language drives sales," *Proceedings of AMCIS-14*.
- James Russell. 1980. "A circumplex model of affect," *Journal of Personality and Social Psychology*, 39(6): 1161-1178.
- Jin Wang, Liang-Chih Yu, K. Robert Lai and Xuejie Zhang. 2016. "Community-based weighted graph model for valence-arousal prediction of affective words," *IEEE/ACM Transaction Audio, Speech and Language Processing*, 24(11): 1957-1968.
- Wen-Li Wei, Chung-Hsien Wu, and Jen-Chun Lin. 2011. "A regression approach to affective rating of Chinese words from ANEW," *Proceedings of ACII-II*, 121-131.
- Liang-Chih Yu, Jin Wang, Robert Lai and Xuejie Zhang. 2015. "Predicting valence-arousal ratings of words using a weighted graph method," *Proceedings of ACL/IJCNLP-15*, 788-793.
- Liang-Chih Yu, Lung-Hao Lee, Shuai Hao, Jin Wang, Yunchao He, Jun Hu, K. Robert Lai, and Xuejie Zhang. 2016a. "Building Chinese affective resources in valence-arousal dimensions," *Proceedings of NAACL/HLT-16*, 540-545.
- Liang-Chih Yu, Lung-Hao Lee and Kam-Fai Wong. 2016b. "Overview of the IALP 2016 shared task on dimensional sentiment analysis for Chinese words", *Proceedings of IALP-16*, 156-160.