

# CKIP Valence-Arousal Predictor for IALP 2016 Shared Task

Hsin-Yang Wang

Institute of Information Science  
Academia Sinica  
Nankang, Taipei, Taiwan  
hywangntut@iis.sinica.edu.tw

Wei-Yun Ma

Institute of Information Science  
Academia Sinica  
Nankang, Taipei, Taiwan  
ma@iis.sinica.edu.tw

**Abstract**—Sentiment analysis is an important task in natural language processing and computational linguistic. Automatic sentiment analysis has been widely applied to opinion reviews and social media for a variety of applications, such as marketing and customer services. The dimensional approach can provide more fine-grained sentiment analysis in which each vocabulary is assigned two continuous numerical values - valence and arousal. Our goal is to predict the both values for the unseen vocabularies. In this paper we propose a combination of three rating predictors - E-HowNet knowledge based, word embedding based and single character based predictors to predict Chinese vocabularies. The final evaluation result shows our approach achieved MAE of 0.583 and PCC of 0.862 on the valence, and MAE of 1.307 and PCC of 0.630 on the arousal.

**Keywords** - Sentiment Analysis; Dimensional Approach; Knowledge Base; Deep Learning; Word Embedding

## I. INTRODUCTION

Sentiment analysis, also known as opinion mining, aims to identify and extract the subjective information from source material. To get the information behind the word, speaker or writer's attitudes which respect to certain topics or overall contextual polarity of a document plays an essential role. A basic task to identify these attitudes is to find out the entity feature (such as word or sentence) in the document is positive or negative. The representation of these features can be generally divided into two different approaches - categorical and dimensional. Categorical approaches represent the features as several discrete classes such as Ekman's six basic emotions (e.g., anger, happiness, fear, sadness, disgust and surprise) [1], while the dimensional approach represents the features as continuous numerical values on multiple dimensions, such as valence-arousal (VA) space [2].

In IALP 2016 shared task [3], the task focuses on the Chinese word feature. Each participant builds their own system to automatically predict the valence-arousal ratings of the Chinese words. The participants could use the training set (CVAW, 1,653 annotated words with valence-arousal ratings) [4] given by the organizers and any other publicly available data to predict the valence-arousal rating of unseen Chinese words. The valence-arousal rating has been defined in real numbers from 1 to 9. The performance of the system is evaluated by examining the difference between system-predicted ratings and human-annotated ratings, using Mean absolute error (MAE) [5] and Pearson correlation coefficient ( $r$ ) [6] as evaluation metrics.

The paper is structured as follows; Section 2 introduces the approach we proposed. Section 3 shows the experiment result. We give our conclusion and future work in section 4.

## II. PROPOSED APPROACH

Our proposed system is based on three rating predictors: 1) *E-HowNet VA predictor*, 2) *Word Embedding VA predictor*, 3) *Single Character Arousal Predictor*. A weighting function is used to control the influence of these three predictors. The system architecture is shown in Figure. 1. In the following sub-sections, we'll talk about the idea of each sub-system, how we implement them, the problems we encountered, and how do we solve these problems.

### 2.1 E-HowNet Valence-Arousal Predictor

Our basic idea to predict the valence-arousal rating is to use the information from our E-HowNet knowledge base. An example of E-HowNet tree structure is shown in Figure. 2. The resource can also be found on our CKIP website [7]. Each word recorded in E-HowNet contains a high quality definition of the word itself and the relations with other words, included synonym, hypernym and hyponym relations.

In this work, we use synonym relation between the given (training) word and target (testing) word to predict the valence-arousal rating. The intuition is that we believe the words in a synset should have similar valence-arousal rating. To make this idea much easier to be understood, in here we use a simple example to explain how we use



Figure 2. Example of E-HowNet tree structure

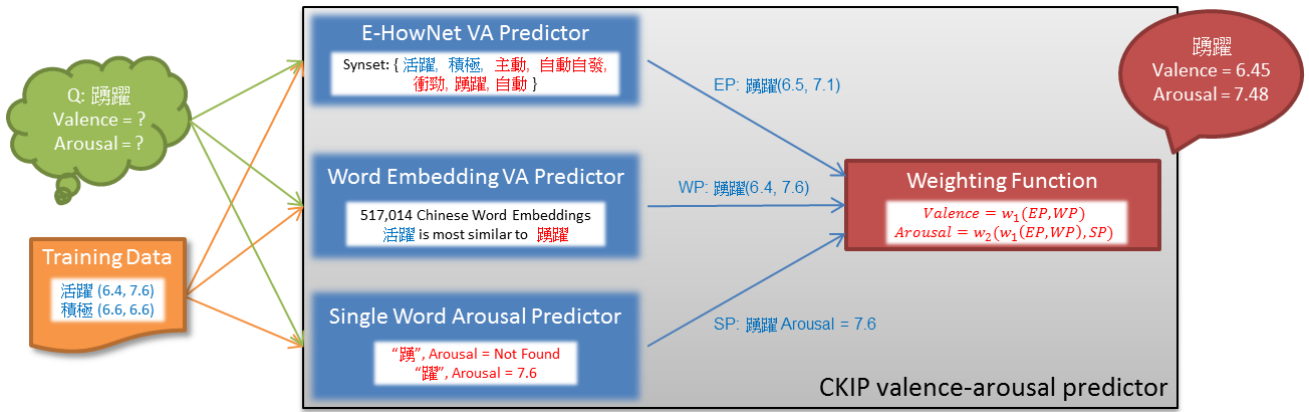


Figure 1. System architecture of CKIP valence-arousal predictor, with a simple example question - “踴躍”

synonym relation to predict valence-arousal rating.

active 積極 [主動, 自動自發, 活躍, 衝勁, 積極, 踴躍, 自動]	
<b>詞彙訊息</b>	
詞彙:	活躍
詞性:	VH14
英文意涵:	enliven, animate, invigorate
定義式:	{active 積極}

Figure 3. Example Synset: {active|積極}

For example, a synset shows in Figure. 3 contains seven synonym words “主動”, “自動自發”, “活躍”, “衝勁”, “積極”, “踴躍”, “自動” that has same E-HowNet expression - {active|積極}. Also, the training data from the organizer has produce the valence-arousal rating of the two words 活躍(6.4, 7.6) and 積極(6.6, 6.6). In this case, our system will predict the valence-arousal rating of the synset as the average rating of given word valence and arousal {active|積極}(6.5, 7.1), then it will spread this rating into the other unscored words in the synset. The result rating will be 主動(6.5, 7.1), 自動自發(6.5, 7.1), 衝勁(6.5, 7.1), 踴躍(6.5, 7.1), 自動(6.5, 7.1).

This idea has demonstrated its effectiveness that can receive high accurate predicting result in our experiment, which we shown the result in section 3.1.

## 2.2 Word Embedding Valence-Arousal Predictor

The previous prediction system shows a reliable result, but it still has low coverage of words. There are several words which cannot be founded in E-HowNet so they are not spread by any given word’s valence-arousal rating. To solve this problem, we construct another prediction method based on word embeddings.

Word Embedding is a practical deep learning technic used in several NLP tasks. Based on a large text corpus, the method could use the context information of the word to represent the words in a continuous vector space. One advantage of the word embedding is that we can use similarity measurement to find some similar words with a specific word. We have applied this advantage into our task to find the similar given (training) words of the target (testing) word.

As the experiment result shows in section 3.2, word embeddings still have problem of word coverage. To handle this issue, we use single-character words to predict embedding of the multi-character words that does not have the word embeddings. That is said, given an uncovered word, we predict its embedding through averaging the embeddings of its characters, which are learned when they are regarded as single-character words in the corpus. e.g.  $e_{\text{癡狂}} = \text{average}(e_{\text{癡}}, e_{\text{狂}})$ .

## 2.3 Single Character Arousal Predictor

The previous two predictors based on the knowledge base and word embeddings get a great success on predicting the valence, but they perform not very well on arousal. To enhance the performance of arousal prediction, we propose a simple but effective idea based on words’ morphological structure. We first predict the arousal rating of each character through averaging the arousal values of all words that contains the character. (For instance, the character – “爽” turns out having higher arousal rating, since there are many multi-character words with higher arousal ratings which contain “爽”, such as “爽快” and “爽朗”. And the character – “平” turns out having lower arousal rating since there are many multi-character words with lower arousal ratings which contain “平”, such as “平淡” and “平凡”.) Once all characters’ arousal ratings are obtained, the arousal rating of a given testing word can then be predicted through averaging the arousal values of its single characters. We show the result in section 3.4.

## III. EXPERIMENT

We use training set given by the organizer with 10-fold cross validation to evaluate our proposed approach. To compare with the current existed model, we use Weighted Graph Model with community [8] as the baseline.

For the E-HowNet knowledge base, we use E-HowNet Ver.2.0, which contains over 90,000 Chinese words in 5,852 synset. For the training of word embeddings, we use CNA+ASBC (中央社語料+中研院平衡語料庫) as our training corpus, the corpus size is 2.41GB. We use CBOW in the word2vec [9] for training, the setting parameters are: dimension is 300, window size is 10, negative sample is 10, and iteration is 15. The output embedding contains 517,014 Chinese words.

### 3.1 E-HowNet Valence-Arousal Predictor

Table 1 shows the E-HowNet VA predictor result. The valence both on MAE and correlation get a better result than the baseline, but the arousal gets poorer result than the baseline. This phenomenon was due to the assumption that the knowledge base can only provide valence information, but lack on the arousal information. And also, about one-third (544 words) cannot be predicted by the E-HowNet VA predictor, we'll improve this in the next section.

TABLE I. E-HOWNET VA REDICTOR RESULT

Method	Valence		Arousal		Unhandled Word
	MAE	r	MAE	r	
Baseline	0.770	0.897	0.613	0.694	-
E-HowNet	0.543	0.905	0.727	0.686	544 / 1,653

### 3.2 Word Embedding Valence-Arousal Predictor

Table 2 shows the result of the Word Embedding VA predictor. By only use the most similar word embeddings to predict valence and arousal (Top 1), the result seems well. But we have noted that the most similar word may have totally opposite valence rating, which could cause a huge error. (For instance, the most similar word of 美夢 is 惡夢) So we use plural similar words and average their valence-arousal rating to predict the target (testing) word. In our experiment, we've used Top 3, 5, 10, and 20. The result shows Top 10 get a great improvement on both MAE and correlation.

TABLE II. WORD EMBEDDING VA PREDICTOR RESULT

Method	Valence		Arousal		Unhandled Word
	MAE	r	MAE	r	
Baseline	0.770	0.897	0.613	0.694	-
Only Embedding <sup>1</sup> (Top 1)	0.736	0.806	0.912	0.532	19 / 1,653
Only Embedding <sup>2</sup> (Top 1)	0.736	0.807	0.912	0.531	0 / 1,653
Only Embedding <sup>2</sup> (Top 10)	0.604	0.887	0.769	0.637	0 / 1,653

1. Only predict the valence-arousal rating of the word that has word embedding  
2. Also use single-character words to predict word embedding

### 3.3 Combination of previous two predictors

On the combination of E-HowNet VA predictor and Word Embedding VA predictor, our primeval strategy is to use the E-HowNet with a higher priority, the word embeddings only predict the words that does not shown in E-HowNet. The idea of this strategy is that E-HowNet VA predictor has shown a better result than the Word Embedding VA predictor. Besides this strategy, we also utilize the information from both knowledge base and word embeddings, using a linear combination and get a great improvement, compared with the primeval strategy. We believe the reason of this improvement is that knowledge base and word embeddings provide different surface of information, and that could help the combined system make a better prediction. We have set the weighting between two models ( $w_1$  shows in Figure. 1.) as 3:7, 5:5, and 7:3. In our experiment, 7:3 and 5:5 gets the

better result, which we submitted as Run-1 and Run-2. The combined system result shows in Table 3.

TABLE III. COMBINED SYSTEM RESULT

Method	Valence		Arousal		Unhandled Word
	MAE	r	MAE	r	
Baseline	0.770	0.897	0.613	0.694	-
E-HowNet	0.543	0.905	0.727	0.686	544 / 1,653
Only Embedding (Top 1)	0.736	0.807	0.912	0.531	0 / 1,653
E-HowNet then Word Embedding (Top 1)	0.632	0.856	0.791	0.625	0 / 1,653
E-HowNet + Word Embdding (Top 1) (7:3)	0.608	0.865	0.770	0.638	0 / 1,653
E-HowNet + Word Embdding (Top 1) (5:5)	0.626	0.859	0.785	0.625	0 / 1,653
Only Embedding (Top 10)	0.604	0.887	0.769	0.637	0 / 1,653
E-HowNet + Word Embedding (Top 10) (7:3)	0.541	0.907	0.714	0.688	0 / 1,653
E-HowNet + Word Embedding (Top 10) (5:5)	0.540	0.911	0.710	0.701	0 / 1,653

### 3.4 Add Single Character Arousal Predictor

Table 4 shows the result after adding the Single Character arousal predictor into our combined system. The both two models have got the improvement on the MAE and correlation of arousal.

TABLE IV. SINGLE CHARACTER AROUSALPREDICTOR RESULT

Method	Valence		Arousal		Unhandled Word
	MAE	r	MAE	r	
Baseline	0.770	0.897	0.613	0.694	-
E-HowNet + Word Embedding (Top 10) (7:3)	0.541	0.907	0.714	0.688	0 / 1,653
E-HowNet + Word Embedding (Top 10) (5:5)	0.540	0.911	0.710	0.701	0 / 1,653
[Run-1] E-HowNet + Word Embedding (Top 10) (7:3) + Single Character	0.541	0.907	0.684	0.740	0 / 1,653
[Run-2] E-HowNet + Word Embedding (Top 10) (5:5) + Single Character	0.540	0.911	0.686	0.744	0 / 1,653

### 3.5 Final Result

Table 5 shows the final result announced by the organizer of IALP 2016 shared task. Our submitted two models - Run-1 and Run-2 has get a great success on the valence rating prediction, which is rank 1 and rank 3 in all the submitted models. The result reflects our assumption on using both E-HowNet knowledge base and word embeddings can get a suitable predicting rating on the

valence. The result of arousal rating prediction is much poorer than our cross validation result shows in Table 4, we believe this was due to a highly difference between the training set and testing set, which we consider to use another strategy on solving this issue.

TABLE V. FINAL RESULT

Method	<i>Valence</i>			<i>Arousal</i>		
	<i>MAE</i>	<i>r</i>	<i>Rank</i>	<i>MAE</i>	<i>r</i>	<i>Rank</i>
[ <i>Run-1</i> ] E-HowNet + Word Embedding (Top 10) (7:3) + Single Character	0.601	0.854	3/32	1.303	0.620	13/32
[ <i>Run-2</i> ] E-HowNet + Word Embedding (Top 10) (5:5) + Single Character	0.583	0.862	1/32	1.307	0.630	14/32

#### IV. CONCLUSION

We have demonstrated our approach, based on three sub-systems with a weighting function. The testing result shows our approach is suitable on predicting the valence-

arousal rating in IALP 2016 shared task. In the future, we will consider more relations in the E-HowNet such as hypernym or hyponym to increase the coverage of words for the E-HowNet VA predictor.

#### REFERENCES

- [1] Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion*, 6:169-200.
- [2] James A. Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.
- [3] [http://nlp.innobic.yzu.edu.tw/tasks/dsa\\_w/index.html](http://nlp.innobic.yzu.edu.tw/tasks/dsa_w/index.html)
- [4] 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 Trans. Audio, Speech and Language Processing*, 24(11):1957-1968..
- [5] [https://en.wikipedia.org/wiki/Mean\\_absolute\\_error](https://en.wikipedia.org/wiki/Mean_absolute_error)
- [6] [https://en.wikipedia.org/wiki/Pearson\\_product-moment\\_correlation\\_coefficient](https://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient)
- [7] <http://ehownet.iis.sinica.edu.tw/index.php>
- [8] Liang-Chih Yu, Jin Wang, K. Robert Lai and Xuejie Zhang. 2015. Predicting valence-arousal ratings of words using a weighted graph method. In *Proc. of ACL/IJCNLP-15*, pages 788-793.
- [9] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, Jeffrey Dean, 2013. Distributed representations of words and phrase and their compositionality, *Advances in neural information processing systems*.