Valence-Arousal Prediction of Chinese Words with Multi-layer Corpora

Xinrui Zhang, Piyuan Lin*, Siyuan Chen, Hongjie Cen, Jundong Wang, Qiangjia Huang, Yuhong Xu, Jiecong Tang, Peijie Huang

College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, Guangdong, China

pyuanlin@scau.edu.cn

***Abstract*—This paper presents our valence-arousal (*VA*) prediction method in the IALP 2016 Shared Task of Dimensional Sentiment Analysis for Chinese Words. Dimensional approach represents affective states as continuous numerical values in multiple dimensions, such as the *VA* space, thus allowing for more fine-grained sentiment analysis. For *VA* prediction, existing works usually selected similar seeds for an unseen word based on semantic similarity by using ontology or word2vec. However the semantic similarity is sometimes quite different from the similarity of valence/arousal. Therefore, this paper proposes a *VA* prediction method with multi-layer corpora to address such difference. In semantic layer, we get the top *N* most similar words by word2vec. Then, we screen the selected similar words based on the training data in *VA* layer. Finally, we use external corpora of affective polarity and intensity lexicons to make further filtering. Experimental results show that the proposed methods in this study to predict the value of *VA* yields good performance for Chinese words, especially in *V* dimension.**

Keywords-multi-layer corpora; valence-arousal (VA) prediction; affective lexicon; word2vec

I. Introduction

With the development of online social network services, sentiment analysis has become a research hot pot for automatically identifying affective information from texts [1-4]. In sentiment analysis, affective states are generally represented using either categorical or dimensional approaches. The categorical approach represents affective states as several discrete classes such as binary (positive and negative) and Ekman's six basic emotions (e.g., anger, happiness, fear, sadness, disgust and surprise) [5]. The dimensional approach has drawn considerable attention in recent years as it can provide a more fine-grained sentiment analysis. It represents affective states as continuous numerical values on multiple dimensions, such as valence-arousal (*VA*) space [6], as shown in Fig. 1.

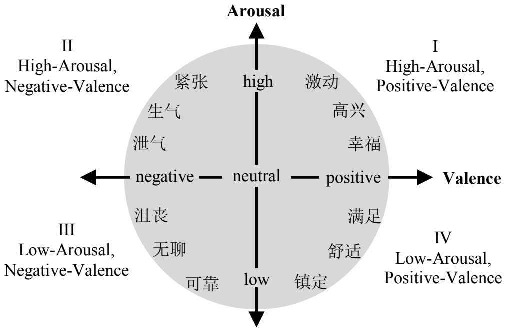

Figure 1. Two-dimensional valence-arousal space.

The valence represents the degree of pleasant and unpleasant (or positive and negative) feelings, and the arousal represents the degree of excitement and calm [7]. Based on such a two dimensional representation, any affective state can be represented as a point in the *VA* coordinate plane by determining the degrees of valence and arousal of given words [8-10].

For *VA* prediction, affective lexicons with *VA* ratings are useful resources but few exist, especially for the Chinese language [7]. Wei et al. [8] manually created an affective lexicon with small number (only 162) of Chinese *VA* words. Recently, Yu et al. [7] built a new lexicon called Chinese valence-arousal words (CVAW) containing 1,653 words by five annotators, which is also used as the training data of this shared task.

VA prediction methods usually start from a set of words with labeled *VA* ratings (called seeds). The *VA* rating of an unseen word is then estimated from semantically similar seeds [11]. Wei et al. [8] trained a linear regression model in a cross-lingual manner using the *VA* ratings of a set of English seed words (source) and their translated Chinese seed words (target) such that the *VA* ratings can be transformed from a source language to a target language. Malandrakis et al. [9] used a kernel function to combine the similarity between seeds and unseen words into a linear regression model. Yu et al. [10] extends the idea of pagerank on a graph in two aspects of *VA* and proposes weighted graph model considers both the relations of multiple nodes and the similarity weights among them which assigns an equal weight to the edges connected between an unseen word and its neighbor nodes. Recently, Wang et al. [11] take into account community-based weighted graph model for *VA* prediction of affective words, and they calculated the similarities of ANEW words using the cosine measure between their corresponding word vectors obtained using the continuous bag-of-words (CBOW) model of word2vec [12].

The major limitation most of the existing works suffer from is that the similar seeds selected for an unseen word by these methods may have quite different ratings of valence/arousal, or even an inverse polarity with the unseen word. In fact, the similarity between two words estimated using ontology or word2vec is sometimes quite different from the similarity of valence/arousal. To address this problem, this paper proposes a *VA* prediction method with multi-layer corpora. In semantic layer, similar to the existing works, we get the top *N* most similar words by word2vec. Then, we screen the selected similar words based on the training data in *VA* layer. Finally, we use external corpora to make further filtering. Particularly, an affective polarity lexicon to remove some

*Corresponding author

words of opposite polarity for *V* prediction, and an affective intensity lexicon for *A* prediction.

The rest of this paper is organized as follows. Section II describes the proposed system for *VA* ratings. Section III summarizes the comparative results of different methods for *VA* prediction. Conclusions are finally drawn in Section IV.

II. PROPOSED METHOD

A. Overall Framework

Fig. 2 shows the overall framework of the proposed method. It is worth notice that, taking into account the differences of *V* and *A* dimensions, we use two different methods for *V* and *A* ranting respectively, and each method uses three-layer-structure separately to screen gradually the closest words set with unseen: semantic layer, *VA* layer and external corpora layer. It performs in the following steps:

1. Semantic layer: for unseen word, getting the similar words set semantically by word2vec.

2. *VA* layer: we use different methods to screen words by *VA* training set (CVAW). For *V* we employ K-Means clustering, and for *A* we utilize improved Manhattan distance.

3. External corpora layer: in this layer, we utilize external corpus to revise the word set for final prediction. An affective polarity lexicon and an affective intensity lexicon are used for *V* and *A* prediction respectively.

B. Semantic Layer

For any unseen word, we firstly get the word set in semantic layer. That is say, these words (seeds) have high similarity semantically with unseen word. We calculate the similarities of all words (train and test words) using the cosine measure between their corresponding word vectors which obtained using 400-dimension continuous vector representations of words, trained on large amounts of Chinese Wikipedia data using the word2vec tool. In this layer, processing of *V* is the same as *A*. After this layer, we can feed top *N* most similar words into next layer for further screening.

C. VA Layer

1) V prediction using K-Means clustering

In this layer, we use 1-dimension of *V* of top *N* words from semantic layer to cluster by K-Means. It is say that we project the *VA* space point into *V* coordinate, and fragment these projection values into *K* groups (clusters). After that, words belongs to the same group is nearest with each other in valence ratings. Then, we calculate each group scores by weights sum all word similarity with the unseen in group, and exclude those in the group with the lowest score for improving evaluation of *V*. This way, we get the top *M* closer words with unseen in *V* rating for next layer.

2) A prediction using improved Manhattan distance

Taking into account the difference between *V* and *A* on the emotional expression, we adopt another method to predict the *A* rating in *VA* layer. For an unseen word w, we use the top *N* words got from the semantic layer to calculate the first *VA* value by weighted similarity sum. The valence of w is denoted as val_w, and computed as the following formula:

$$val_{\mathrm{w}} = \frac{\sum_{\mathrm{w}_i \in \mathrm{S(w)}} \mathrm{sim}(w, w_j) \cdot V_{\mathrm{w}_i}}{\sum_{\mathrm{w}_i \in \mathrm{S(w)}} \mathrm{sim}(w, w_i)}, \qquad (1)$$

where $S(w)$ is the set of the top N similar words of w. V_{w_i} is the V value of word w_i in training data. $sim(w, w_i)$ represents the similarity between words w and w_i. The arousal of w, denoted as aro_{w}, is computed similar to val_{w}.

In *VA* layer, we use the improved distance measurement method to reduce the error of the prediction of the model in first step. By cross validation, we find that the Manhattan distance is better than the Euclidean distance. The improved Manhattan distance of the unseen word w and word w_j in training data is computed as follows:

$$MD_j = \alpha \cdot \left| V_{\mathrm{w}_j} - val_{\mathrm{w}} \right| + \left| A_{\mathrm{w}_j} - aro_{\mathrm{w}} \right|, \mathrm{j} = 1, \ldots, \mathrm{n}, \quad (2)$$

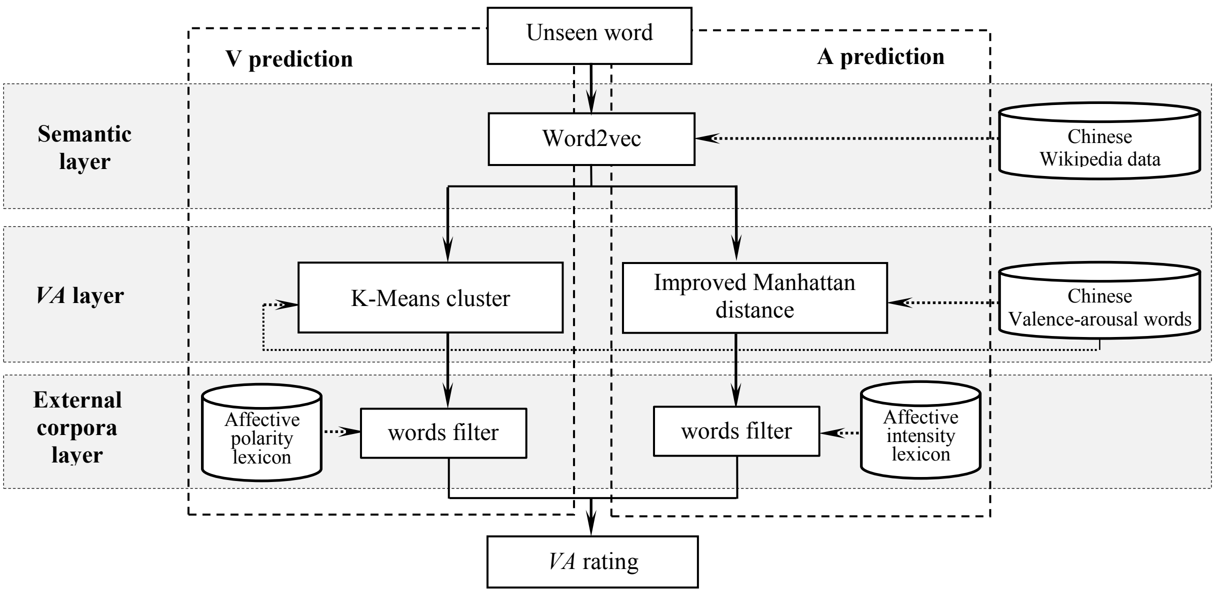

Figure 2. Overall framework of the proposed method.

where V_{w_j} and A_{w_j} are the V and A values of word w_j in training data, and n is the size of training data. α is a factor for adjusting the weight of A and V in computing the Manhattan distance (in our experiment, α is set as 4 after cross validation). Then, we choose M words with smallest Manhattan distance to the unseen word w for A rating of w for next layer.

D. *External Corpora Layer*

1) *V filtering using affective polarity lexicon*

The valence represents the degree of pleasant and unpleasant feelings, so we use affective polarity lexicon to clear up some noisy words after semantic layer and *VA* layer. In each word of top *M* similar words, if the candidate word is opposite to the positive/negative of unseen word by looking up an affective lexicon originates in NTUSD[1] and HowNet[2], it would be removed. By this revise, we can remove a few noise words from the candidate words and predict value of *V* by weighted cumulating the remaining candidate words using formula (1). This method is really efficient and the parameters were chosen via a grid search on the training set.

2) *A filtering using affective intensity lexicon*

Because of the arousal represents the degree of excitement and calm, we cannot use the same lexicon as *V*. So, we use the affective intensity lexicon provide by irlab of Dalian University of Technology[3] to evaluate the *M* similar words from the second step and remove potential noise words with opposite to the intensity of the unseen word, remaining *T* words to estimate a value of *A*:

$$A_{predict} = \frac{1}{T} * \sum_{j=1}^{T} A_j, \tag{3}$$

III. EMPIRICAL EVALUATION

A. *Task*

Due to the limited availability of *VA* lexicons, especially for Chinese, the objective of the task of Dimensional Sentiment Analysis for Chinese Words which is the shared task organized in the 20th International Conference on Asian Language Processing (IALP 2016) is to automatically acquire the valence-arousal ratings of Chinese affective words. Given a word, participants are asked to provide a real-valued score from 1 to 9 for both valence and arousal dimensions, indicating the degree from most negative to most positive for valence, and from most calm to most excited for arousal. The training data of CVAW consists of 1,653 affective words annotated with *VA* ratings annotated by five annotators. And the test data for shared task contains 1,149 testing instances.

B. *Metrics*

The performance is evaluated by examining the difference between machine-predicted ratings and human-annotated ratings (*V* and *A* are treated independently). The evaluation metrics include Mean absolute error (MAE) and Pearson correlation coefficient (PCC). MAE and PCC defined as follow:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |A_i - P_i|, \tag{4}$$

$$\mathrm{PCC} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{A_i - \bar{A}}{\sigma_A}\right)\left(\frac{P_i - \bar{P}}{\sigma_P}\right), \tag{5}$$

where A_i denotes the human-annotated ratings, P_i denotes the machine-predicted ratings, n is the number of test samples, A and P respectively denote the arithmetic mean of A and P, and σ is the standard deviation.

C. *Model Selection*

In this task, we tried a variety of methods to predict *VA* rating, including weighted graph (weight represents similarity of words or Euclidean distance of words in *VA* space) community algorithm, word2vec, KNN and K-Means. In addition, we also have a combination of these methods for verification. 5 fold cross-validation results of different methods for *V* and *A* are shown in Table I and Table II. "Corpus" stands for affective polarity lexicon in *V* prediction and affective intensity lexicon in *A* prediction. ED and MD stand for Manhattan distance and Euclidean distance, respectively. SD stands for similarity distance.

TABLE I. COMPARISON OF DIFFERENT METHODS FOR V.

Methods	V_MAE	V_PCC
Community(SD)	1.225	0.527
Community(ED)	1.055	0.623
KNN+Community(SD)	0.826	0.786
KNN+Community(SD)+Corpus	0.669	0.851
Word2vec	0.814	0.805
Word2vec+K-Means	0.795	0.802
Word2vec+Corpus	0.629	0.883
Word2vec+K-Means+Corpus	**0.627**	**0.884**

TABLE II. COMPARISON OF DIFFERENT METHODS FOR A.

Methods	A_MAE	A_PCC
Community(ED)	0.992	0.345
Community(SD)	0.963	0.363
Word2vec	0.826	0.578
Word2vec+ED	0.823	0.589
Word2vec+MD	0.824	0.586
Word2vec+ED+Corpus	0.825	0.578
Word2vec+MD+Corpus	**0.820**	**0.593**

As shown in Table I and Table II, for both *V* and *A*, the methods with multi-layer corpora gets better performance. The "Word2vec+K-Means+Corpus" and the "Word2vec+ MD+corpus" methods get the best verification results in *V* and *A* prediction, respectively.

D. *Evaluation Results*

The shared task attracted 22 research teams to participate, 16 teams submitted their system results. For formal testing, each participant has a right to submit at

---

1 www.datatang.com/data/44317
2 www.datatang.com/data/46754
3 www.datatang.com/data/45448

most two runs. Organizers also provide a similarity-based regression model as a baseline for performance reference. In total, official organizer get 32 runs. For each dimension, they rank MAE and PCC independently and calculate the mean rank (average of MAE rank and PCC rank) for ordering system performance.

1) Two runs of our methods

Based on cross-validation, two runs of our methods submitted to the shared task final test are as follows:

Run1: This run is the "full" proposed method using "Word2vec+K-Means+Corpus" for *V* prediction and "Word2vec+ MD+Corpus" for *A* prediction.

Run2: Because the performance gained of K-Means and intensity lexicon is limited in the cross verification of the prediction of *V* and *A*, we submit the second run using "Word2vec+Corpus" for *V* prediction and "Word2vec+ MD" for *A* prediction.

2) DSAW final test

Table III and Table IV show the evaluation results *V* and *A* of the final test respectively. "Our" is the better one of the two runs submitted by us. It is slightly surprised that in the two runs of our method, Run2 outperforms Run1, which may indicates less effect of K-Means and affective intensity lexicon for the prediction of *V* and *A* respectively in the data set of the shared task. The "Best" indicate the high score of mean rank achieved in DSAW task. The "Baseline" represents the baseline of the official run. "Rank of teams" is sorted by mean rank of MAE and PCC of the best run of each team.

TABLE III. V EVALUATION RESULTS OF DSAW FINAL TEST

	Valence MAE (rank of runs)	**Valence PCC (rank of runs)**	**Rank of teams**
Our	0.768 (15)	0.865 (1)	4
Best	0.583 (4)	0.862 (3)	1
Average	0.807 (-)	0.763 (-)	-
Baseline	1.407 (31)	0.674 (28)	16

TABLE IV. A EVALUATION RESULTS OF DSAW FINAL TEST

	Arousal MAE (rank of runs)	**Arousal PCC (rank of runs)**	**Rank of teams**
Our	1.305 (21)	0.604 (12)	12
Best	1.212 (8)	0.671 (1)	1
Average	1.261 (-)	0.469 (-)	-
Baseline	1.567 (32)	0.473 (19)	16

From Table III we can know that we got the best evaluation result about "Valence PCC", and for "Valence MAE", we also got a good result. Table IV shows that our prediction of *A* is still some distance from the best.

IV. CONCLUSION AND FUTURE WORK

This paper proposes the *VA* prediction method with multi-layer corpora from team of South China Agricultural University (SCAU) that participated in the IALP 2016 Shared Task of Dimensional Sentiment Analysis for Chinese Words. By using different kinds of corpora, word vectors trained on Chinese Wikipedia data using the word2vec tool provides candidate similar words in semantic level, *VA* training set (CVAW) gives valuable screening of the selected similar words in *VA* level, and finally external corpora of affective polarity and intensity lexicons makes further filtering. Experiments on training set show that the proposed method yielded good cross validation result.

In final test, the performance of *V* rating of our method is better than that of *A* rating, which is mainly because that the similar semantic of words calculated by word2vec is not fit to represent similarity of *A* dimension, and there is a lack of good affective intensity lexicon. So, future work will focus on the improvement of *A* prediction and extend the *VA* prediction from the word-level to the sentence- and document-levels.

ACKNOWLEDGMENT

The research work was supported by National Natural Science Foundation of China under Grant No. 71472068, and the Special Funds for the Cultivation of Guangdong College Students' Scientific and Technological Innovation under Grant No. pdjh2016b0087.

REFERENCE

[1] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and trends in information retrieval, vol. 2, no. 1-2, 2008, pp.1-135.

[2] R. A. Calvo and S. D'Mello, "Affect detection: An interdisciplinary review of models, methods, and their applications," IEEE Trans. Affective Computing, vol. 1, no. 1, 2010, pp.18-37.

[3] B. Liu, "Sentiment analysis and opinion mining," Synthesis Lectures Human Lang. Technol., vol. 5, no. 1, 2012, pp. 1-167.

[4] R. Feldman, "Techniques and applications for sentiment analysis," Communications of the ACM, vol. 56, no. 4, 2013, pp.82-89.

[5] P. Ekman. An argument for basic emotions. Cognition and Emotion, vol. 6, no. 3-4, 1992, pp.169-200.

[6] J. A. Russell, "A circumplex model of affect," Journal of personality and social psychology, , vol. 39, no. 6, 1980, pp.1161-1178.

[7] L. C. Yu, L. H. Lee, S. Hao, et al., "Building Chinese affective resources in valence-arousal dimensions," Proceedings the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 16),pp.540-545.

[8] W. L. Wei, C. H. Wu, and J. C. Lin, "A regression approach to affective rating of Chinese words from ANEW," Proceedings the 2011 International Conference on Affective Computing and Intelligent Interaction (ACII-11), pp. 121-131.

[9] N. Malandrakis, A. Potamianos, I. Elias, et al., "Kernel models for affective lexicon creation," Proceedings of the 12th Annual Conference of the International Speech Communication Association (INTERSPEECH 11), 2011, pp. 2977-2980.

[10] L. C. Yu, J. Wang, K. R. Lai, et al., "Predicting valence-arousal ratings of words using a weighted graph method," Proceedings the 53rd Annual Meeting of the Association for Computational Linguistics and 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP 15), pp. 788-793.

[11] J. Wang, L. C. Yu, K. R. Lai et al., "Community-based weighted graph model for valence-arousal prediction of affective words", IEEE/ACM Trans. Audio, Speech and Language Processing, vol. 24, no. 11, pp. 1957-1968, 2016.

[12] T. Mikolov, K. Chen, G. Corrado, et al., "Efficient estimation of word representations in vector space," Proceedings the 1th International Conference on Learning Representations (ICLR 2013), 2013, pp.1-12