LDCCNLP at IJCNLP-2017 Task 2: Dimensional Sentiment Analysis for Chinese Phrases Using Machine Learning

Peng Zhong[1], **Jingbin Wang**[1]
[1] College of Mathematics and Computer Science, Fuzhou University, Fuzhou, P.R. China

Abstract

Sentiment analysis on Chinese text has intensively studied. The basic task for related research is to construct an affective lexicon and thereby predict emotional scores of different levels. However, finite lexicon resources make it difficult to effectively and automatically distinguish between various types of sentiment information in Chinese texts. This IJCNLP2017-Task2 competition seeks to automatically calculate Valence and Arousal ratings within the hierarchies of vocabulary and phrases in Chinese. We introduce a regression methodology to automatically recognize continuous emotional values, and incorporate a word embedding technique. In our system, the MAE predictive values of Valence and Arousal were 0.811 and 0.996, respectively, for the sentiment dimension prediction of words in Chinese. In phrase prediction, the corresponding results were 0.822 and 0.489, ranking sixth among all teams.

1 Introduction

Emotional analysis is a technique of mining and identifying potential emotional information in texts. Such techniques can allow for the automatic analysis of public opinion to help guide government policy making, can help firms improve products and services in response to customer feedback, and can help improve medical treatment through automatically identifying emotional labels in patients' medical records.

In general, affective states can be described in two ways: categorical representation and dimensional representation. (Calvo and D'Mello, 2010). Many studies have examined sentiment classification (Jiang et al., 2011; Boni et al., 2015). The categorical analysis represents affective states as several discrete class, such as positive and negative (Ekman, 1992). However, the categorical representation can't express the fine-grained intensity of emotion. Therefore, dimensional representation has emerged as an important topic in sentiment analysis emerge for application in different fields (Li and Hovy, 2014; Preotiuc-Pietro, 2015; Choudhury, 2012; Wang et al., 2016a; 2016b; Yu et al., 2016a). The dimensional representation can represent any affective state as a point in a continuous multi-dimensional space. In linguistic theory, sentiment can be represented as a point in a bi-dimensional space defined in terms of Valence (the degree of pleasant and unpleasant) and Arousal (the degree of exciting and calm) (Russell, 1980). The resulting space is called VA space.

Dimensional sentiment analysis has attracted widespread attention for tasks involving natural language processing. The dimensional affective lexicon has been widely used in dimensional sentiment analysis, such as in CVAW (Yu et al., 2016b), ANEW (Bradley and Lang, 1999) and so on. Given the limited availability of affective lexicons, especially for Chinese, the aim of the IJCNLP 2017 task 2 is to automatic identify VA ratings of word-level and phrase-level in Chinese. This paper presents a system that uses word embeddings (Mikolov et al., 2013a; 2013b; Pennington et al., 2014; Yu et al., 2017) to represent the Chinese word and phase as

Proceedings of the 8th International Joint Conference on Natural Language Processing, Shared Tasks, pages 84–88,
Taipei, Taiwan, November 27 – December 1, 2017.

input, and uses the regression method to fit the valence and arousal ratings (Brereton and Lloyd, 2010).

The rest of this paper is organized as follows. Section 2 presents the method used to train the word vectors of the traditional Chinese corpus. Section 3 describes the evaluation methods and results. Section 4 presents conclusions.

2 PROPOSED METHOD

2.1 Data Collection

The task of this competition is based on the prediction of traditional Chinese texts. The domain and dimension of the corpus will affect the quality of the word vector (Lai et al., 2016). Therefore，for this task, we collected two kinds of corpus for the construction of a traditional Chinese corpus:

1. Chinese Wikipedia Dumps

The Wikipedia Extractor tool is used to extract articles from Wikipedia text. After removing punctuation, OpenCC is used to convert all remaining text into Traditional Chinese.

2. Taiwan Commercial Times

To obtain a richer traditional Chinese text, we used crawler programs to obtain news content on the webpages of the Taiwan Commercial Times. text is obtained with the scale of 0.65G in total after cleaning up punctuations and non-Chinese content.

After obtaining the traditional Chinese corpus, the two pieces of corpora are merged, scaling 1.6G. We then used the CKIP word segmentation system to segment the text to produce a traditional Chinese corpus containing 767,103 words.

2.2 Word Vector Training

After constructing the Chinese corpus, we needed to transform the words into numerical vectors usable by the machine learning regression algorithm. Word embedding techniques can represent words as continuous low-dimensional vectors which contain the semantic and syntactic information of words. The semantic similarity between words can be obtained by calculating the distance between vectors. We use two methods to train two types of word vectors.

Word embeddings can be obtained by using a neural network to train the language model (Xu and Rudnicky, 2000). Using the relationship between word contexts, we can obtain the feature output of the hidden layer in the process of word prediction. Google's Word2Vec tool is based on this principle, and we can use it to train different dimensions of the word vector.

Words that always appear together are semantically similar, and their meaning may be reflected by co-occurrence context (Chen and You, 2002). Through matrix co-occurrence, we can also train low-dimensional word vectors with semantic similarity. We use the Glove tool (Pennington et al., 2014) to train this kind of co-occurrence based word vector. For each type of word vector, we will train five dimension vectors, with dimensions of 100, 150, 200, 250 and 300.

2.3 Model

For each subtask, we use five regression models in machine learning to carry out experiment: Ridge Regression, K Nearest Neighbors (KNN), Decision Trees(DT), Support Vector Regression (SVR), and AdaBoost. These models are implemented by scikit-learn.

3 Experiments and Evaluation

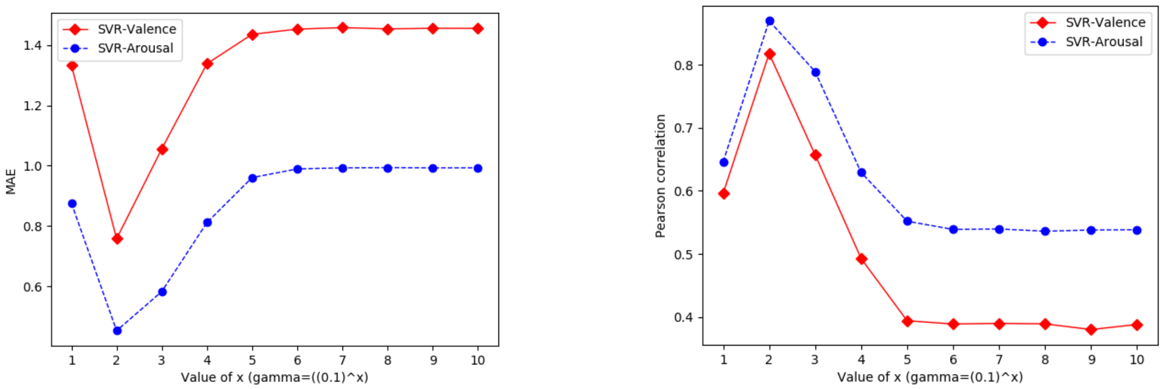
3.1 Dataset

The Shared Task2 contains two subtasks: namely VA rating prediction of words and phrases in Chinese.

The training set of the word VA ratings prediction task 2,802 emotion words with valence-arousal ratings in the CVAW2.0 affective lexicons (Yu et al., 2016) provided by the organizer. In the prediction task of the VA rating of the phrases, the training set contains 2250 phrases, VA ratings also annotated. The contest test set consists of 750 words and 750 phrases that are not annotated with VA ratings.

3.2 Evaluation metrics.

Performance was evaluated by comparing the difference between the predicted values and the corresponding actual values in the test set. The IJCNLP 2017 Task 4 published results for all participants using both mean absolute error (MAE) and Pearson correlation coefficient (PCC).



(a) (b)

Figure 1: Parameter gamma selection for support vector regression methods, evaluated on the development set of Chinese phase using MAE and PCC.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|A_i - P_i\right|$$

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

where A_i is the actual value, P_i is the predicted value, n is the number of test samples, $\bar{A}$ and $\bar{P}$ respectively denote the arithmetic mean of A and P, and σ is the standard deviation. The MAE measures the error rate and the Pearson correlation coefficient shows the correlation between the actual value and predicted result.

3.3 Implementation details

For the two sub-tasks, we divided the experiment into two parts.

In the word prediction task, we use word vectors of different dimensions and different types obtained from training mentioned above to convert the words in the training set into corresponding word vectors. An n-dimensional random vector is generated as corresponding token with the value for each dimension within the range of -0.25 and +0.25, given that the word doesn't belong to the realm of our corpus, where n represents current word embedding dimension.

For the phrase predicting task, such steps are executed, follow these steps: First, with CKIP system phrases are divided into words, with a maximum of three words per phrase. Then we translate each word in phrases into corresponding word vector, the conversion method used here identical to the transformation process of the word task; finally, each phrase is embodied by phrase vectors of 3*n with feature vectors of all words preserved; for a phrase comprised of less than 3 words, 0 is used to fill the vacant values of phrase vectors.

Before each experiment, we disrupted the training set and performed 5-fold validation to adjust and record the parameters of each model based on the cross-validation results.

Different model parameters affect the performance of the regression algorithm. In each subtask, we adjust the model parameters based on the performance of forecast results with evaluation matrice.

The following is an example of parameter adjustment process for the SRV algorithm with phrase VA prediction task based on the 300-dimensional word vector produced by the Glove, in which we use the MAE and Pearson correlation coefficients to recognize the prediction performance. First, the kernel function of appropriate SRV algorithm is selected by predicting the result of crossing validation sets. After the comparison, we use rbf kernel function. Then, the value of the gamma parameter of the rbf kernel is adjusted again by the result of 5-fold cross validation.

In Fig. 1(a), the point x on the abscissa is an integer ranging from 1 to 10, representing the change in the value of gamma, and the value of gamma is equal to the x power of 0.1. The ordinate is the MAE evaluation result from different gamma values with the SVR algorithm. It can be seen from Fig. 1 that when x = 2, that is, gamma 0.01, the model has the best MAE value for Valence and Arousal prediction of the phrase.

Figure 1(b) shows the influence of different gamma values over the Pearson correlation coefficients of the predicted results. When gamma is 0.01, the coefficient value for Valence and Arousal prediction results reaches highest. Results.

For the two sub-tasks, we use the Word2Vec and GloVe word vector with 100, 150, 200, 250, and 300 dimensions to compare the effect of the regression model. The following figure shows the results of the phrase prediction task.

Figure 2 evaluates the prediction results of the

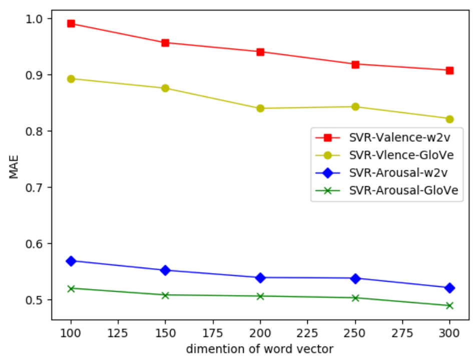


Figure 2: Selection of the dimension and type of word vector.

phrase test set using the SVR regression model. The abscissa represents the number of word vector dimensions from 100 to 300. For each dimension, we perform experiments on the word vector trained by Word2Vec (Mikolov et al., 2013a; Mikolov et al., 2013b) and GloVe (Pennington et al., 2014). The ordinate is the MAE evaluation value of the test set prediction results for specific model.

1) As the number of word vector dimensions increases, the predictive result is getting better, both for the Word2Vec word vector and the GloVe word vector, the same outcome seen for richer features of word vectors. Experiments show that the same conclusion can be drawn from comparing the predictions of words and phrases in the other proposed regression models.

2) The word vectors based on word concurrence matrix trained by GloVe is better than that of Word2Vec related to word contexts in VA rating prediction task for phrases. However, in word pre-

Word	Valence		Arousal	
	MAE	PCC	MAE	PCC
KNN	1.219	0.521	1.235	0.346
DT	1.372	0.407	1.385	0.181
SVR	**0.811**	**0.769**	**0.996**	**0.479**
AdaBoot	0.934	0.696	1.091	0.364
Ridge	0.857	0.754	1.056	0.451
Phase	Valence		Arousal	
	MAE	PCC	MAE	PCC
KNN	0.916	0.632	0.605	0.743
DT	1.092	0.495	0.821	0.543
SVR	**0.822**	**0.762**	**0.489**	**0.828**
AdaBoot	1.035	0.723	0.622	0.744
Ridge	1.779	0.322	1.019	0.481

Table 1: The best predictions for each model in word and phrase prediction tasks

diction tasks, using Word2Vec to train the word vector performs better than GloVe.

Table 1 summarizes the best predictive results for each model in the two subtasks. The best results for predictions of words and phrases are obtained using the SVR regression models.

4 Conclusions

To automatically identify the valence-arousal ratings for lexicon augmentation, this paper presents a machine learning regression model to predict valence-arousal ratings of Chinese words and phrases. We use the word embeddings to produce the word vector for the lexicon vocabulary. Experiments on both Chinese words and phrases show that the proposed method provides good predictive results. The SVR method outperformed other regression methods. Future work will focus on further improving predicative performance. The modifier can be considered as the characteristic of Chinese phase for the predictive result.

References

Mohammad Al Boni, Keira Zhou, Hongning Wang, and Matthew S. Gerber. 2015. Model adaptation for personalized opinion analysis. In *Proceedings of the 53rd*

Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL/IJCNLP-15), pages 769–774.

Margaret M. Bradley and Peter J. Lang. 1999. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1, Univerity of Florida, Gainesville, FL.

R. G. Brereton and G. R. Lloyd. 2010. Support vector machines for classification and regression. *Analyst*, 135(2):230–267.

R. A. Calvo and Sidney. D'Mello. 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Trans. Affective Computing*, 1(1): 18-37.

Keh Jiann Chen and Jia Ming You. 2002. A study on word similarity using context vector models. *Computational Linguistics and Chinese Language Processing*, 7(2):37–58.

Munmun De Choudhury, Scott Counts, and Michael Gamon. 2012. Not all moods are created equal! Exploring human emotional states in social media. In *Sixth international AAAI conference on weblogs and social media*, pages 66–73.

Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion*, 6(3-4), 169-200.

Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. 2011. Target-dependent twitter sentiment classification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-11)*, pages 151–160.

Siwei Lai, Kang Liu, Shizhu He, and Jun Zhao. 2016. How to generate a good word embedding. *IEEE Intelligent Systems*, 31(6):5–14.

Jiwei Li and Eduard Hovy. 2014. Sentiment Analysis on the People's Daily. In *Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP-14)*, pages 467–476.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *Computer Science*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013b. Distributed representations of words and phrases and their compositionality. *Advances in Neural Information Processing Systems*, 26:3111–3119.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proc. of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP-14)*, pages 1532–1543.

Daniel Preotiuc-Pietro, Johannes Eichstaedt, Gregory Park, Maarten Sap, Laura Smith, Victoria Tobolsky, H. Andrew Schwartz, and Lyle Ungar. 2015. The role of personality, age, and gender in tweeting about mental illness. In *Proc. of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal To Clinical Reality*, pages 21–30.

James A 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. 2016a. Community-based weighted graph model for valence-arousal prediction of affective words. *IEEE/ACM Transactions on Audio Speech and Language Processing*, 24(11):1957–1968.

Jin Wang, Liang-Chih Yu, K. Robert Lai, and Xuejie Zhang. 2016b. Dimensional sentiment analysis using a regional CNN-LSTM model. In *Proc. of the 54th Annual Meeting of the Association for Computational Linguistics (ACL-16)*, pages 225-230, 2016.

Wei Xu, Alexander I Rudnicky. 2000. Can artificial neural networks learn language models? In *The Proceedings of the Sixth International Conference on Spoken Language Processing*, pages 202–205.

Liang-Chih Yu, Lung-Hao Lee, and Kam-Fai Wong. 2016a. Overview of the IALP 2016 shared task on dimensional sentiment analysis for Chinese words, in *Proc. of the 20th International Conference on Asian Language Processing (IALP-16)*, pages 156-160.

Liang-Chih Yu, Lung-Hao Lee, Shuai Hao, Jin Wang, Yunchao He, Jun Hu, K. Robert Lai, and Xuejie Zhang. 2016b. Building Chinese affective resources in valence-arousal dimensions. In *Proceedings of the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL/HLT-16),* pages 540-545.

Liang-Chih Yu, Jin Wang, K. Robert Lai, Xue-jie Zhang. 2017. Refining word embeddings for sentiment analysis. In *Proc. of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP-17)*, pages 545–550.