

previous studies (Wei et al., 2011; Gökçay et al., 2012; Malandrakis et al., 2013; Paltoglou et al., 2013; Yu et al., 2015). In this study, we follow the manual annotation approach to build a Chinese affective lexicon and corpus in the VA dimensions.

3 Affective Resource Construction

This section describes the process of building Chinese affective resources with valence-arousal ratings, including the CVAW and CAVT.

The CVAW is built on the Chinese affective lexicon C-LIWC, and then annotated with VA ratings for each word. Five annotators were trained to rate each word in the valence and arousal dimensions using the Self Assessment Manikin (SAM) model (Lang, 1980). The SAM model provides affective pictures, which can help annotators in determining more precise labels when rating the words. The valence dimension uses a nine degree scale. Values 1 and 9 respectively denote the most negative and positive degrees of affect. Point 5 means a neutral emotion without specific tendency. The arousal dimension uses a similar scale to denote calm and excitement. Using this approach, each affective word can be annotated with VA ratings (determined by the average rating values provided by the annotators) to form the CVAW.

To build the CVAT, we first collected 720 web texts from six different categories: news articles, political discussion forums, car discussion forums, hotel reviews, book reviews, and laptop reviews. A total of 2,009 sentences containing the greatest number of affective words found in the C-LIWC lexicon were selected for VA rating. The Google app engine was then used to implement a crowdsourcing annotation platform using the SAM annotation scheme. Volunteer annotators were

asked to rate individual sentences from 1 to 9 in terms of valence and arousal. Each sentence was rated by at least 10 annotations. Once the rating process was finished, a corpus cleanup procedure was performed to remove outlier ratings and improper sentences (e.g., those containing abusive or vulgar language). The outlier ratings were identified if they did not fall into the interval of the mean plus/minus 1.5 standard deviations. They were then excluded from the calculation of the average VA ratings for each sentence.

4 Results

4.1 Analysis Results of CVAW

A total of 1,653 words along with the annotated VA ratings were included in the CVAW lexicon, yielding the (mean, standard deviation) = (4.49, 1.81) for valence and (5.48, 1.26) for arousal. To analyze differences between the annotations, we compared the VA values rated by each annotator against their corresponding means across the five annotators to calculate the error rates using the following metrics.

- *Mean Absolute Error (MAE):*

$$MAE = \frac{1}{n} \sum_{i=1}^n |A_i - \bar{A}_i|,$$

- *Root Mean Square Error (RMSE):*

$$RMSE = \sqrt{\sum_{i=1}^n (A_i - \bar{A}_i)^2 / n},$$

where A_i denotes the valence or arousal value of word i rated by an annotator, \bar{A}_i denotes the mean valence or arousal of word i calculated over the five annotators, and n is the total number of words in the CVAW.

	MAE		RMSE	
	Valence	Arousal	Valence	Arousal
Annotator A	0.4934	1.3479	0.6372	1.6411
Annotator B	0.5972	0.7821	0.7488	0.9929
Annotator C	0.5817	1.1393	0.7423	1.4302
Annotator D	0.5188	0.8226	0.6614	1.0374
Annotator E	0.6258	1.0200	0.7970	1.2700
(Mean, SD)	(0.56, 0.05)	(1.02, 0.21)	(0.72, 0.06)	(1.27, 0.24)

Table 1: Analysis of error rates of different annotators for building the Chinese VA lexicon.