Figure 2: Scatter plot of the CVAW lexicon.
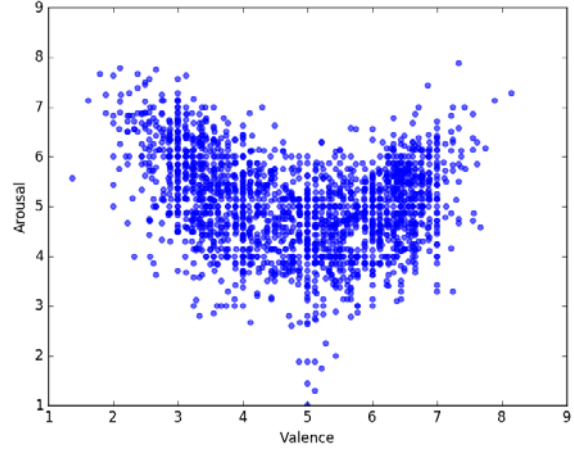

Figure 3: Scatter plot of the CVAT corpus.

	Num. of texts	Num. of tokens	Avg. tokens	Valence			Arousal		
				MAE	RMSE	r	MAE	RMSE	r
ANEW vs Forum	20	15,035	751.75	1.20	1.55	0.77	0.72	0.85	0.27
CVAW vs CVAT	2,009	70,456	35.07	1.20	1.52	0.54	1.01	1.28	0.16
Book Review	287(14%)	8,217	28.63	1.00	1.31	0.41	**0.89**	**1.11**	0.21
Car Forum	257 (13%)	12,261	47.71	1.48	1.77	0.30	0.92	1.15	0.10
Laptop Review	183 (9%)	5,374	29.37	**0.95**	**1.21**	0.55	1.07	1.40	0.04
Hotel Review	301 (15%)	7,268	24.15	1.35	1.73	0.59	0.93	1.17	**0.22**
News Article	542(27%)	21,923	40.45	1.11	1.40	**0.61**	1.11	1.40	0.17
Politics Forum	439 (22%)	15,413	35.11	1.28	1.61	0.51	1.04	1.32	0.19

Table 2: Results of using the CVAW lexicon to predict the VA ratings of the CVAT corpus.

Table 1 shows the error rates of the annotators in rating the VA values of words in the CVAW. Overall, for all metrics the error rates of arousal ratings were greater than those of valence ratings. In addition, the annotators produced more consistent error rates (around 0.49~0.63 for MAE and 0.64~0.80 for RMSE) in the valence dimension than those (around 0.78~1.35 for MAE and 0.99~1.64 for RMSE) in the arousal dimension. These findings indicate that the degree of arousal was more difficult to distinguish than valence.

Figure 2 shows a scatter plot of words in the CVAW, where each point represents the mean of the VA values as rated by the annotators. Several words (translated from Chinese) were marked in the VA space for reference, e.g., *victory* (7.8, 7.2), *trust* (7.8, 5.8), *pain* (2.4, 6.8), *kill* (1.6, 7.8), *tedi-* *ous* (3.4, 3), *fault* (3.6, 4.6), *agree* (6.4, 4.4) and *relaxed* (6.2, 2.0).

4.2 Analysis Results of CVAT

A total of 2,009 sentences with VA ratings were included in the CVAT corpus, yielding the (mean, standard deviation) = (4.83, 1.37) for valence and (5.05, 0.95) for arousal. The distribution of the six categories and their word counts in CVAT are shown in Table 2. The largest category was News (27%), while the smallest one was Laptop (9%). Figure 3 shows a scatter plot of VA ratings for all sentences in CVAT. It is similar with the plot of the CVAW, indicating that annotators followed similar guidelines for rating affective words and sentences.

543