SiSP: Japanese Situation-dependent Sentiment Polarity Dictionary

Anonymous ACL submission

Abstract

In order to deal with the variety of meanings and contexts of words, we created a Japanese Situation-dependent Sentiment Polarity Dictionary (SiSP) of sentiment values labeled for 20 different situations. This dictionary was annotated by crowdworkers with 25,520 Japanese words, and consists of 10 responses for each situation of each word. Using our SiSP, we predicted the polarity of each word in the dictionary and that of dictionary words in sentences considering the context. In both experiments, situation-dependent prediction showed superior results in determining emotional polarity.

1 Introduction

Understanding human emotions from facial images, voice, texts, and so on is becoming very important both in academia and industry. Emotion polarity dictionaries are used to analyze emotions from texts. Most of the existing emotion polarity dictionaries are based on a single word labeled as positive or negative, or they only classify words into a number of class categories. However, even a single word can have many different meanings and give different impression when used in different contexts and situations. For example, the word *fast* can be positive when it means that a racing car is fast, but it can have a negative meaning when you are walking with a friend and you want to complain that your friend is walking too fast. Many current emotion polarity dictionaries have only a single label and cannot handle such a variety of situations and meanings. Meanwhile, emotion polarity dictionaries that consider various categories are annotated only with class labels and ignore the strength of the emotion polarity of words in the category.

In this study, we developed a Situation-dependent Sentiment Polarity Dictionary (SiSP) with individual numerical labels for 20 different situations. To the best of our knowledge, SiSP is the first situation-dependent sentiment polarity dictionary. We will make it an open source upon acceptance. In addition, we have demonstrated the baseline performance of the polarity prediction of words in two senarios: that of individual word and that with context.

2 Related works

2.1 Sentiment lexicon

Most sentiment lexicons are lists of words labeled in a positive or negative direction. They are often created manually because of the subjective nature of sentiment labels. Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2001) is a dictionary of over 6,000 words classified into 125 categories. This dictionary has been used to extract political sentiments from tweets and to predict the onset of depression from SNS text.

The Affective Norms for English Words (ANEW) lexicon (Bradley and Lang, 1999) consists of 1,024 English words labeled from 1 to 9 in terms of the Valence-Arousal-Dominance (VAD) model. SentiWordNet (Esuli and Sebastiani, 2006) (Baccianella et al., 2010) is an extension of WordNet (Miller, 1995) that scores words on a scale of 0.0 to 1.0 for positive, negative, and neutral, and is normalized so that the sum of each category score is 1. SentiWordNet is also labeled in a semi-supervised manner. Many words are classified as neutral, with no polarity and a very high level of noise.

The SiSP created in this study has a numerical value from 0 to 1 for each of the 20 different situations with labels of positive, negative, neutral (between positive and negative), irrelevant (the word has nothing to do the situation), and unintelligible.

2.2 Named Entity Recognition

Named Entity Recognition (NER) is a task to extract unique expressions contained in sentences. It extracts Named Entities from sentences and classifies them into proper nouns such as names of people, organizations, and places, and predefined expressions such as dates, time expressions, quantities, and amounts. For these expressions, a distinction is made between *between* (B) for the first one and *inside* (I) for the second one. Tokens that do not belong to any entity are assigned *outside* (O). This distinction is called BIO notation. For example, in the sentence ‘Mark Watney visited Mars’, if the person tag is ‘Person’ and the location tag is ‘Location’, Mark is a B-Person, Watney is an I-Person, visited is an O because it does not belong to any token. Some tasks classify place names into detailed locations such as cities, states, countries, etc., and some tasks set subcategories such as politicians or celebrities for person names. Typical datasets are CoNLL2002 and CoNLL2003 (Sang and De Meulder, 2003). These datasets assign four types of tags (Person, Organization, Location, Miscellaneous including all other types of entities) to sentences consisting of newspaper articles in Spanish, Dutch, English, and German. There are other datasets for other languages as well (Singh, 2008) (Shaalan, 2014) (Benikova et al., 2014), (Piskorski et al., 2017).

3 SiSP Construction

Existing emotion polarity dictionaries cannot reflect the situation the words are used in. Therefore, in this study, we have created a new emotion polarity dictionary with 20 different situations. The 20 situations we use are as follows (see Table 1): Economy, Communication, Parenting, Health, Will Motivation, Employment, Personal Relations, Family, Tension, Mental State, Work, Going Out, Home, Busyness, Motivation, Sleep, Appetite, Curiosity, Physical State, and Love. The dictionary assumes a wide range of situations that can evoke emotions. We used 25,520 words from Takamura’s (Takamura et al., 2005) and Kobayashi’s (Kobayashi et al., 2005) dictionaries (14,400 and 11,120 words, respectively). Five label options were provided: positive, negative, neutral, irrelevant, and unintelligible. The annotations were done via a crowdsourcing service, and 10 people were asked to label each word and situation. We set two very strict check problems to guarantee the quality of the dictionary, in which only 40% of the participants survived. In summary, the SiSP labels were labeled by 10 people (who passed the check problems) for 20 situations. Therefore, the labels do not simply indicate whether each word is positive or negative for each situation, but rather indicate the degree of each option.

Table 2 shows a few examples of words in SiSP that have different labels depending on the situation. The word ‘Interference’ has (Positive, Negative) = (0.1, 0.6) in *love*, while it changes to (0.5, 0.2) in *mental*. The word ‘sold-out’ has a positive value of 0.8 in *economy*, while it has a positive value of 0.1 in *going out* The word ’hostility’ has a negative value of 0.9 in *personal relations*, while it changes to 0.3 in *parenting*.

4 Experiments

The dictionary we created can be used to classify the polarity predictions of the words in a particular situation. We performed prediction in two different experimental settings. The first task was to predict the polarity of individual words. The second task was to predict the polarity of words contained in sentences. In this paper, the results will be presented only for *mental state* due to the page limitation. Please see our appendix for detailed experimental results. As a baseline where a situation is not considered, the 20 situation-dependent polarity labels were majority voted and a single polarity score was assined to each word. By comparing the results with this baseline, the significance of the situational labeling in this study is discussed.

The original SiSP labels were based on 10 responses for each word and each of the 20 situation. In this experiment, neutral, irrelevant, and unintelligible, are treated as one label O. Then, the label with the largest vote among positive, negative, and O is dealt as the ground truth class. The training data contain 1,118 positives, 3,364 negatives, and 9,624 O’s for *mental state* situations.

4.1 Polarity prediction of individual words (words alone)

The words in the dictionary are divided into training data and test data. We used word2vec (Mikolov et al., 2013) to vectorize the words in the dictionary, and Suzuki’s pre-trained 300-dimensional model was used as the model of word2vec. We used this vector as input and classify the polarity of words by using a support vector machine (SVM).

The classification results of the test data are presented on the left side of Table 3. The results of precision and recall for positives were remarkably

Table 1: SiSP data

# of words	25,520				
# of evaluators/word	10				
Labels	Positive	Negative	Neutral	Irrelevant	Unintelligible
Situations	Economy	Communication	Parenting	Health	Will Motivation
	Employment	Personal Relations	Family	Tension	Mental State
	Work	Going Out	Home	Busyness	Motivation
	Sleep	Appetite	Curiosity	Physical State	Love

Table 2: Example of SiSP words

Words	Situation	Labels				
		Positive	Negative	Neutral	Irrelevant	Unintelligible
Interference	Love	0.1	0.6	0.2	0.1	0
	Mental State	0.5	0.2	0	0.3	0
Sold-out	Economy	0.8	0	0.1	0.1	0
	Going Out	0.1	0.1	0.3	0.5	0
Hostility	Personal Relations	0	0.9	0	0.1	0
	Parenting	0	0.3	0.1	0.6	0

low, which may be due to the unbalance of the dataset. As for the baseline, none of the items were judged as positive, and most of them were classified as O. As for the F1 score of negative, the situation-dependent prediction is very much improved from the baseline (from 0.191 to 0.558).

4.2 Polarity prediction of words in a sentence (words with context)

Next, we conducted an experiment to see if the performance can be improved by adding contextual information. The task is based on the NER method, where sentences are morphologically analyzed and labels are assigned to each word. Words that are not in the dictionary are also treated as O. In this study, we collected 20,530 tweets using the Twitter API, and used those tweets that contained at least one word registered in SiSP as input data. The task is to predict the polarity of the word in SiSP that appears in only one tweet out of all the tweets collected as the test data. Therefore, the test data include the SiSP word that appears only once in all the tweets collected, and the rest are used as the training data. The words to be predicted are not included in the training data. At the test time, the model would predict the polarity of completely unknown words. As a model, we used BERT's (Devlin et al., 2018) Japanese pre-training model. The results are shown on the right side of Table 3. Comparing the *mental state* and baseline, the precision, recall, and F1 score were better for the *mental state* for both positive and negative classes. Particularly, the F1 score and recall for the *mental state* were significantly better than that for baseline in the negative class. It is also shown that the context of the sentence generally helps the model to predict the polarity more accurately when we compare the results in Table 3 (note that the test data are different from each other).

5 Discussion

Regarding the polarity classification for individual word alone, the reason for the poor accuracy is that there are so many words with labels of O (Pos: 1,118, Neg: 3,364, and O: 9,624). Therefore, we used up-sampling and down-sampling so that number of words with polarity labels are the same as those with O labels. Up-sampling for word2vec was done by simply repeating the training data. Down-sampling for NER task was done by eliminating sentences that consist only of words labeled O.

The results of up-sampling are shown on the left side of Table 4. We can observe a significant improvement in the values of precision, recall, and F1 for polarity labels in the *mental state*.

It is apparent that the up-sampling result outperforms the left side of Table 4 related to mental state of precision, recall, and F1-score for positive and negative labels. In particular, there is a remarkable improvement in the positive label. Precision for

Table 3: Results of polarity classification for *mental state* situations

	Words alone				Words with context			
	Precision	**Recall**	**F1**	**Accuracy**	**Precision**	**Recall**	**F1**	**Accuracy**
Pos-baseline	0.000	-	-	0.860	0.354	0.374	0.364	0.867
Pos-mental state	0.026	0.655	0.050	0.795	0.387	0.431	0.408	0.825
Neg-baseline	0.116	0.545	0.191	0.856	0.394	0.355	0.373	0.834
Neg-mental state	0.498	0.633	0.558	0.832	0.556	0.570	0.563	0.803
O-baseline	0.981	0.722	0.832	0.717	0.806	0.814	0.810	0.710
O-mental state	0.898	0.640	0.450	0.755	0.744	0.719	0.731	0.663

Table 4: The effects of up- and down-sampling of the dataset for polarity classification for mental state situations

	Up-sampling for words alone				Down-sampling for words with context			
	Precision	**Recall**	**F1**	**Accuracy**	**Precision**	**Recall**	**F1**	**Accuracy**
Pos	0.444	0.475	0.459	0.767	0.383	0.453	0.415	0.832
Neg	0.696	0.462	0.555	0.788	0.613	0.568	0.590	0.806
O	0.593	0.688	0.637	0.603	0.735	0.729	0.732	0.668

Table 5: Prediction performance for mental state when the prediction is done word by word

	Precision	**Recall**	**F1**	**Accuracy**
Pos	0.314	0.497	0.385	0.844
Neg	0.289	0.624	0.395	0.798
O	0.865	0.670	0.755	0.655

up-sampling mental state was better than that for no up-sampling in positive by 0.418 (from 0.026 to 0.444). In negative label, precision for up-sampling mental state was the best score of all settings.

The results of the down-sampling are shown on the right side of Table 4. Compared to the results in Table 3, the F1 score for both negative and positive words showed improvement. Precision for negative words and recall for positive words were also improved. However, down-sampling for words with context does not provide as large an effective as up-sampling for words alone. This may be due to the fact that only O sentences were excluded from the training data, so the words between positive and negative were not balanced.

In addition, to confirm that the model is getting hints from the words around the unknown word, we input the test data word by word instead of tweet by tweet to BERT. The results shown in Table 5 can confirm that the precision and F1 scores for positive and negative are larger when each tweet is input to the model. In particular, precision for the negative class for the *mental state* was better by 0.267.

6 Conclusion

In this study, we proposed the Situation-dependent Sentiment Polarity Dictionary (SiSP), which is a dictionary that considers 20 types of situations, because the existing emotion polarity dictionaries with uni-dimensional positive-negative labels cannot grasp the fact that the polarity can change depending on the context and situation. In the words alone prediction, the results for negative's precision, recall, and F1-score, in particular, were significantly better than the baseline. Additionally, the prediction performance increased by up-sampling the training dataset. We also confirmed that situation-dependent prediction yielded better precision, recall, and F1 score. In polarity prediction of unknown words using tweets as training data, the situation-dependent prediction also showed superior results. The comparison between tweet-by-tweet and word-by-word demonstrated that contextual information in the input data is also important for inferring emotional polarity.

This is the first open source sentiment polarity dictionary with 20 situations with some baseline performance. Our future work includes extending this dictionary to other languages.

7 Acknowledgements

This research has been approved by the ethical committee of the authors' organization. (ID is omitted for double-blind review.)

References

Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani. 2010. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec*, volume 10, pages 2200–2204.

Darina Benikova, Chris Biemann, and Marc Reznicek. 2014. Nosta-d named entity annotation for german: Guidelines and dataset. In *LREC*, pages 2524–2531.

Margaret M Bradley and Peter J Lang. 1999. Affective norms for english words (anew): Instruction manual and affective ratings. Technical report, (Tech. Report C-1). Gainesville: University of Florida, Center for Research in Psychophysiology.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Andrea Esuli and Fabrizio Sebastiani. 2006. Sentiwordnet: A publicly available lexical resource for opinion mining. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*.

Nozomi Kobayashi, Kentaro Inui, Yuji Matsumoto, and Kenji Tateishi. 2005. Collecting evaluative expressions for opinion extraction. *Journal of Natural Language Processing*, 12(3):203–222.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.

James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001.

Jakub Piskorski, Lidia Pivovarova, Jan Šnajder, Josef Steinberger, and Roman Yangarber. 2017. The first cross-lingual challenge on recognition, normalization, and matching of named entities in slavic languages. In *Proceedings of the 6th Workshop on Balto-Slavic Natural Language Processing*, pages 76–85.

Erik F Sang and Fien De Meulder. 2003. Introduction to the conll-2003 shared task: Language-independent named entity recognition. *arXiv preprint cs/0306050*.

Khaled Shaalan. 2014. A survey of arabic named entity recognition and classification. *Computational Linguistics*, 40(2):469–510.

Anil Kumar Singh. 2008. Named entity recognition for south and south east asian languages: taking stock. In *Proceedings of the IJCNLP-08 Workshop on Named Entity Recognition for South and South East Asian Languages*.

Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientations of words using spin model. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, pages 133–140.

A Appendix

Table 6 shows the results for all the other 19 situations that could not be included in the main paper. The separation of training data and test data is the same as that for the mental state.

Table 6: Results of polarity classification for all situations

		Words alone				Words with context			
		Precision	**Recall**	**F1**	**Accuracy**	**Precision**	**Recall**	**F1**	**Accuracy**
baseline	Pos	0.284	0.398	0.331	0.840	0.387	0.431	0.408	0.825
	Neg	0.641	0.420	0.508	0.817	0.556	0.570	0.563	0.803
	O	0.746	0.787	0.766	0.674	0.744	0.719	0.731	0.663
Economics	Pos	0.402	0.445	0.423	0.779	0.416	0.372	0.393	0.814
	Neg	0.718	0.486	0.579	0.792	0.379	0.351	0.365	0.792
	O	0.615	0.704	0.656	0.614	0.717	0.748	0.732	0.634
Communication	Pos	0.444	0.475	0.459	0.767	0.433	0.330	0.374	0.768
	Neg	0.696	0.462	0.555	0.788	0.600	0.549	0.573	0.815
	O	0.593	0.688	0.637	0.603	0.650	0.731	0.688	0.627
Parenting	Pos	0.396	0.428	0.411	0.789	0.335	0.395	0.362	0.834
	Neg	0.624	0.421	0.411	0.789	0.387	0.325	0.353	0.757
	O	0.645	0.721	0.681	0.611	0.721	0.734	0.728	0.628
Health	Pos	0.326	0.393	0.356	0.796	0.318	0.407	0.357	0.824
	Neg	0.666	0.456	0.356	0.808	0.387	0.338	0.361	0.768
	O	0.689	0.745	0.716	0.641	0.721	0.712	0.717	0.614
Will Motivation	Pos	0.321	403	0.357	0.785	0.255	0.463	0.329	0.856
	Neg	0.655	0.403	0.357	0.785	0.406	0.345	0.373	0.819
	O	0.664	0.727	0.694	0.613	0.816	0.775	0.795	0.693
Employment	Pos	0.420	0.475	0.446	0.781	0.459	0.421	0.439	0.822
	Neg	0.723	0.503	0.593	0.801	0.459	0.563	0.479	0.790
	O	0.620	0.695	0.655	0.616	0.751	0.698	0.723	0.646
Personal Relations	Pos	0.450	504	0.475	0.792	0.440	0.594	0.506	0.873
	Neg	0.714	0.470	0.567	0.820	0.440	0.600	0.510	0.830
	O	0.641	0.727	0.681	0.683	0.853	0.749	0.798	0.718
Family	Pos	0.405	0.431	0.418	0.841	0.405	0.431	0.418	0.841
	Neg	0.359	0.355	0.357	0.775	0.359	0.355	0.357	0.775
	O	0.748	0.741	0.745	0.648	0.748	0.741	0.745	0.648
Tension	Pos	0.362	0.319	0.339	0.814	0.362	0.319	0.339	0.814
	Neg	0.485	0.435	0.459	0.811	0.485	0.435	0.459	0.811
	O	0.721	0.760	0.740	0.644	0.721	0.760	0.740	0.644
Mental State	Pos	0.462	0.486	0.474	0.787	0.387	0.431	0.408	0.825
	Neg	0.752	0.518	0.614	0.801	0.556	0.570	0.563	0.803
	O	0.611	0.716	0.659	0.632	0.744	0.719	0.731	0.663
Work	Pos	0.434	0.509	0.469	0.779	0.393	0.444	0.417	0.828
	Neg	0.753	0.522	0.617	0.795	0.449	0.588	0.509	0.798
	O	0.597	0.675	0.633	0.615	0.782	0.699	0.738	0.661
Going Out	Pos	0.288	0.427	0.344	0.829	0.238	0.296	0.264	0.807
	Neg	0.608	0.412	0.491	0.802	0.405	0.416	0.410	0.824
	O	0.730	0.757	0.743	0.653	0.782	0.748	0.765	0.661
Home	Pos	0.341	0.415	0.374	0.809	0.256	0.433	0.321	0.861
	Neg	0.658	0.425	0.516	0.787	0.402	0.353	0.376	0.800
	O	0.666	0.739	0.700	0.624	0.793	0.759	0.775	0.669
Busyness	Pos	0.260	0.335	0.293	0.808	0.273	0.218	0.243	0.970
	Neg	0.587	0.337	0.428	0.795	0.782	0.789	0.786	0.961
	O	0.694	0.761	0.726	0.625	0.962	0.966	0.964	0.936

		word2vec				BERT NER			
		Precision	**Recall**	**F1**	**Accuracy**	**Precision**	**Recall**	**F1**	**Accuracy**
Motivation	Pos	0.412	0.441	0.426	0.776	0.354	0.495	0.413	0.851
	Neg	0.716	0.469	0.567	0.791	0.404	0.387	0.395	0.774
	O	0.602	0.703	0.648	0.604	0.757	0.721	0.738	0.641
Sleep	Pos	0.228	0.351	0.276	0.853	0.271	0.395	0.321	0.892
	Neg	0.625	0.400	0.488	0.810	0.362	0.411	0.385	0.840
	O	0.755	0.796	0.775	0.679	0.862	0.814	0.837	0.743
Appetite	Pos	0.225	0.403	0.289	0.837	0.358	0.421	0.387	0.847
	Neg	0.568	0.383	0.457	0.816	0.318	0.397	0.353	0.837
	O	0.769	0.770	0.769	0.670	0.827	0.776	0.801	0.702
Curiosity	Pos	0.374	0.420	0.395	0.772	0.363	0.467	0.408	0.837
	Neg	0.641	0.399	0.492	0.802	0.422	0.588	0.491	0.834
	O	0.644	0.720	0.680	0.605	0.827	0.729	0.775	0.686
Physical State	Pos	0.353	0.419	0.383	0.794	0.328	0.474	0.388	0.853
	Neg	0.688	0.472	0.560	0.794	0.423	0.500	0.458	0.835
	O	0.675	0.735	0.704	0.636	0.838	0.761	0.798	0.706
Love	Pos	0.284	0.398	0.331	0.840	0.350	0.399	0.373	0.826
	Neg	0.640	0.420	0.508	0.817	0.373	0.373	0.373	0.818
	O	0.746	0.787	0.766	0.674	0.772	0.753	0.762	0.659