Emotion Evaluator: Expanding the Affective Lexicon with Neural Network Model

Junho Lee[1]* **Jaeseo Lim**[1]* **Jooyong Park**[1,2] **Cheongtag Kim**[1,2]
[1]Interdisciplinary Program in Cognitive Science, [2]Department of Psychology
Seoul National University, Republic of Korea
{smbslt3, jaeseolim, jooypark, ctkim}@snu.ac.kr

Abstract

Measuring the emotion in words is valuable in that it analyzes emotions through language. However, it is difficult to find such measurements in low-resource languages. In this paper, we proposed a method to expand the affective lexicon by utilizing the context of words. The proposed model predicted the Valence and Arousal values of words using their dictionary definitions. In Study 1, we reviewed previous studies about the Korean affective lexicon and integrated data from these studies. The model was trained to minimize the MSE error between the Valence and Arousal values of the words and their predictions. We then checked the distribution of Valence and Arousal values of Korean vocabulary by applying our model to the Korean dictionary. In Study 2, a new affective lexicon was built to empirically validate our model. We found a negatively biased error pattern on model predictions and discussed why it happened.

Keywords: Affective Lexicon; Emotion word; Valence; Arousal

Introduction

Psychologists have attempted to measure and classify human emotions for a long time ago. In particular, psycholinguists have considered that the use of language may reflect various psychological processes, including emotions. From this point of view, research has been conducted to evaluate the emotions of words to find the structure of emotions in words used in daily life. (Clore, Ortony, & Foss, 1987; Bradley & Lang, 1999). However, the research on emotions using languages had clear limitations. One limitation was that emotions have been evaluated only for a limited number of words because humans cannot evaluate all the emotions of countless words. Another limitation was that words were evaluated without incorporating their meaning in the rating process. Therefore, the interpretation of the contextual meaning of the words might not be consistent among the people.

Unlike in the case of English, where tens of thousands of emotion word data are accumulated, it is difficult to find a study on the emotion measurement of Korean words. In addition, unlike studies on English words using consistent metrics, studies on Korean words are very difficult to integrate results because of their own metrics. Moreover, since most of the research on Korean words have been conducted by referring to the research on English words, the limitations mentioned above still exist in the evaluation process.

In this study, we proposed a method that automatically evaluated the emotions of Korean words by leveraging the results of previous studies. In Study 1, we built and optimized a neural network model that inferred emotion values from the meaning of words based on the latest natural language processing techniques and definitions of words in the dictionary. In this way, we evaluated the emotion values of all words that existed in the Korean dictionary. In Study 2, while limiting the contextual meaning of words, we discussed the differences between the human ratings and model predictions.

There are three contributions of the present study. The first is to overcome the limitations of research on the Korean affective lexicon by integrating the data of three previous studies. The second is to resolve the ambiguity of previous studies in rating the emotion of words by presenting the context of words as a dictionary definition. The third is that we built a model to evaluate the emotion of words with state-of-the-art techniques, and empirically validated the model by comparing its results with the human ratings, eventually figured out the reason behind the prediction error of the model.

Background

Measurement of emotion

There are two main perspectives on how to measure human emotions. One is the discrete emotion model that explains human emotions can be classified into several basic emotions (Ekman, 1992; Plutchik, 2001). The other perspective is the dimensional emotion model that claims human emotions are continuous values in several dimensions. Russell and Mehrabian's three-factor theory of emotions is a representative study of this perspective (Russell & Mehrabian, 1977). Russell and Mehrabian (1977) defined those human emotions exist in three dimensions: Valence, Arousal, and Dominance (VAD). Each dimension is defined as a bipolar axis of pleasant-unpleasant (V), activation-deactivation (A), and dominance-submissive (D). A lot of research has been conducted especially on the V and A dimensions (the dimension of V and the dimension of A) regarding the measurement of emotion. For example, it has been widely used to measure the emotional intensity of facial expressions, pictures, and words (Adolph & Alpers, 2010; Cuthbert, Schupp, Bradley, Birbaumer, & Lang, 2000; Bradley & Lang, 1999), and was also confirmed to be a significant measurement criterion in behavioral, physiological, and neurological levels (Reuderink, Mühl, & Poel, 2013; Anders, Lotze, Erb, Grodd, & Birbaumer, 2004; Dolcos, LaBar, & Cabeza, 2004).

*Equal contribution.

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

Affective Lexicons and Prediction Modeling

What the word means in the VAD dimension has been investigated for a long time. These emotion word lists have been called various terms such as *Affective Lexicon* (Ortony, Clore, & Foss, 1987), *Affective Norms* (Bradley & Lang, 1999), and *Affective Word List* (Võ, Jacobs, & Conrad, 2006). Research has also been conducted in various languages, such as French (Monnier & Syssau, 2014), Greek (Palogiannidi, Koutsakis, Losif, & Potamianos, 2016), Chinese (Yu et al., 2016)) as well as English.

However, all of these studies have a common limitation that human ratings cannot be performed on numerous already existing words and infinitely emerging new words. Accordingly, recent studies have attempted to overcome this limitation by incorporating methodology in the field of Natural Language Processing. These studies used statistical analysis of words (Recchia & Louwerse, 2015; Vankrunkelsven, Verheyen, De Deyne, & Storms, 2015) or word embedding method (Wang, Yu, Lai, & Zhang, 2016; Sedoc, Preoţiuc-Pietro, & Ungar, 2017) to expand the existing affective lexicon. These studies, however, also have limitations in that the latest deep-learning techniques have not been applied such as Transformer (Vaswani et al., 2017). In addition, they have evaluated the performance of the model within the data they already had on their hand, without collecting new data. Even if cross-validation or bootstrapping has been applied to their model, these methods often cannot prevent overfitting.

Research on Korean Affective Lexicon

Among the many studies that measured the emotions of Korean words, there were three studies that applied the concept of dimensional theory of emotion by Russell and Mehrabian (1977): Park and Min (2005), Rhee and Ko (2013), and Hong, Nam, and Lee (2016). While Russell proposed Valence, Arousal, and Dominance (VAD) as three dimensions of measuring emotions (Russell & Mehrabian, 1977; Russell, 1978), results of research on D dimension were not consistently observed compared to V and A in subsequent studies (Russell, 1983; Watson & Tellegen, 1985). Therefore, the above three studies borrowed only the V and A dimensions showing consistent result and then selected their own new dimensions to complement the D dimension (Table 1).

Park and Min (2005) selected a list of emotion words based on a lexicon about the frequency of Korean words in daily life. As a result, a list of emotion words that were frequently used and can represent various emotions was completed (*Affective Lexicon 1*, hereinafter AL_1). The V dimension can be interpreted as a continuous meaning ($+3 \sim -3$) from positive (pleasant, +3) to negative (unpleasant, -3) as shown in Figure 1. It confirmed that more than half of the words of AL_1 are distributed in the negative area. This is in line with the evolutionary view that emotions have developed to help survival, and negative emotions were beneficial in survival by evading dangerous situations (Vaish, Grossmann, & Woodward, 2008; Lazarus, 2021). Thus, there are more negative emotional expressions from negative emotions than positive emotions. Therefore, AL_1 can be regarded as data that well reflects our emotional structure.

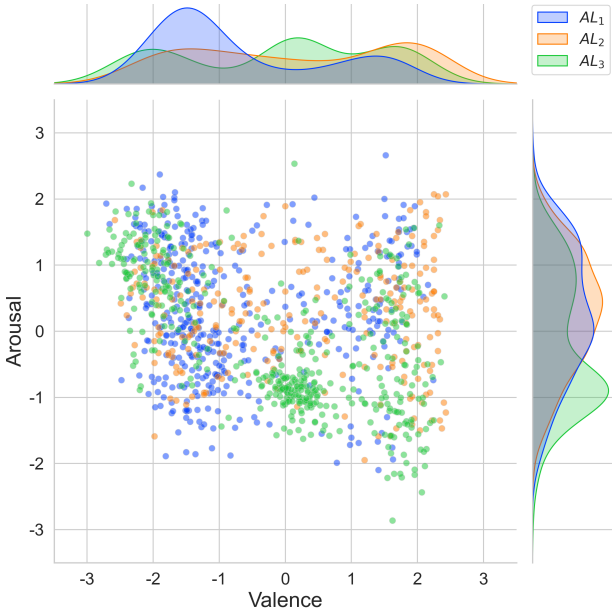

Figure 1: The distribution of each word in three research on Korean Affective Lexicon in Valence-Arousal dimension

Table 1: Summary of research on korean affective lexicon

Dataset	Research	Measurements	# of words
AL_1	Park and Min (2005)	Prototypicality, Familiarity, **Valence, Activation(Arousal)**	434
AL_2	Rhee and Ko (2013)	**Valence, Activation(Arousal)**	267
AL_3	Hong et al. (2016)	Frequency, Concreteness, **Valence, Activation(Arousal)**	450

Rhee and Ko (2013) collected the words on Facebook and excluded words that overlapped with AL_1, then measured their V and A values (hereinafter AL_2). The authors remarked that the distribution of V and A values of this study was not significantly different from that of Park and Min (2005), but in the V dimension, the word distribution of AL_2 was biased in the positive direction compared to the distribution of AL_1. It seems to be due to the origin of words in AL_2. In general, on social media, people write more positive posts than negative ones. Thus, the words in AL_2 were somewhat positively biased compared to the words in AL_1.

Hong et al. (2016) selected some words from the word list from Park and Min (2005) and B. Kim et al. (2010) and con-

structed a Korean affective lexicon by adding neutral words (i.e., objects) unrelated to emotions (hereinafter referred to as AL_3). As a result of the experiment, unlike the previous two studies, specific word clusters appeared in the V-A dimension at 0 and -1, respectively. All words around this point were neutral words that have been newly added such as *건물building, 건전지battery, 비누soap, 볼펜pen*. Therefore, it confirmed that neutral words showed a measure of 0 in the V dimension, referring to pleasant-unpleasant emotions, while -1 in the A dimension, which indicates activation-deactivation of emotion.

Finally, each of the three studies did not show a U-shaped curve in the V-A dimension, unlike studies conducted in other languages. This is presumably because three studies used a very limited number of affective words (unlike large-scale affective lexicon that could cover a wide variety of meanings, such as 1,061 terms in Bradley and Lang (1999) or 20,000 terms in Mohammad (2018) (Table 1). However, after combining the data from the three studies, the distribution of V-A values of words showed a U-shaped curve as the number of words increased including words from various contexts and emotions (Figure 1). Therefore, by synthesizing the data of these three studies, present study expanded the study of Korean emotion words beyond the limitations of individual studies. Accordingly, data incorporating the results of the three studies (hereinafter referred to as AL_t) were used to build the proposed model.

Study 1

In Study 1, we built a model to evaluate the emotion of words with V-A scores. We integrated and prepossessed data, and optimized the model. We then evaluated the V-A values of numerous headwords in the Korean dictionary with the model, and discussed the distribution of emotions of the Korean vocabulary set.

Preprocessing

We set and preprocess the dataset for training the model. The AL_t dataset was built by integrating AL_1, AL_2, and AL_3. Each V and A values were normalized to have the range of -3 to 3. AL_2 included buzzwords (meme), onomatopoeia, and emoticons that were widely used on social media at the time of the study. Among these data, words no longer used or whose meaning cannot be found in the dictionary were excluded from the analysis ($n = 31$). Since AL_3 was built after two other studies, some words in AL_3 were overlapped with the previous two studies (AL_1-AL_3: 32, AL_2-AL_3: 1). The V and A values of the overlapped words were averaged for each word, and the source of the words was set to AL_3. After preprocessing, the definitions of each word were found in the dictionary. Naver dictionary[1] was used to find the definition of words. It provided three definitions from three different sources for each word: *Standard Korean Language Dictionary, Urimalsaem(Korean open dictionary), Korea University Korean Dictionary*(National Institute of Korean Language, 2008, 2016; Research Institute of Korean Studies, 2009). In the process of describing the meaning of the word, definitions from all three dictionaries were considered so that the meaning of the word can be described as abundantly as possible, and the model can learn various tokens during the training process. When there were multiple meanings in a word, the most common meaning was chosen. Homonyms were estimated by considering V and A values because it was not possible to confirm which meaning of the word was presented to the participants during the experiment in the previous studies.

[1] https://dict.naver.com/

Optimization and Test

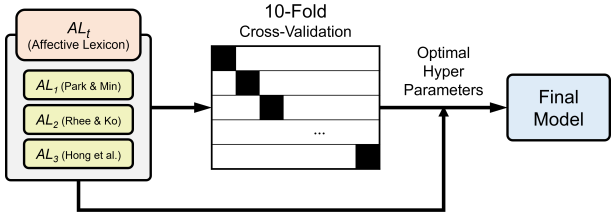

Figure 2: The procedure of model optimization.

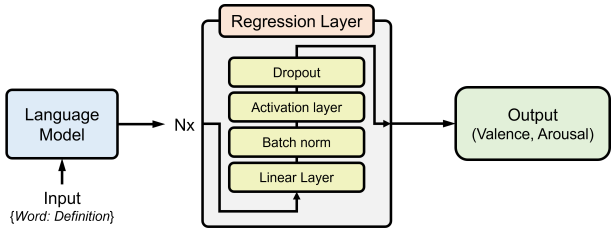

Figure 3: The structure of proposed model

To build a model that predicted V-A values using words and dictionary definitions as inputs, a model with a structure of Figure 3 was proposed. The preprocessed data were used as an input to the model in the form of *word: definition*. The language models were BERT (Devlin, Chang, Lee, & Toutanova, 2018) and ELERTRA (Clark, Luong, Le, & Manning, 2020), which were Korean versions with 40,000 tokens (K. Kim, 2020). To fine-tune the pre-trained language model as much as possible, the word itself was also used as an input. The input text was passed to the language model and transformed into the embedding vectors. The vector of `[CLS]` token of the language model passed through N regression layers, and it became two-dimensional a vector, `(V, A)`. The model was trained to minimize the mean squared error (*MSE*) between the output values and the original V-A values of the input word. To make up for the lack of training data, the optimization process of the hyperparameters and the model structures was conducted with 10-fold cross-validation, and the training/validation set was split into maintaining the proportion of the source of words. The options to search for the best hyperparameter and model structure are shown in Table 2.

Table 2: The tested options to find optimal model. RL means Regression Layer of the model

Hyperparameter	Search Space	Optimal Value
Language model	BERT, ELECTRA	ELECTRA
Batch size	16, 32	32
Learning rate	5e-6, 1e-6, 2e-6, 3e-6, 5e-5, 1e-5, 2e-5, 3e-5	2e-5
Dropout rate	0.2, 0.5	0.5
Activation Function	ReLU, GELU, LeakyReLU	LeakyReLU
Optimizer	Adam, AdamW	AdamW
# of RL	1, 3, 5	3
Dimension of RL	192, 384, 768, 1536, 3072	768

An additional model which consisted of a combination of Word2Vec and LightGBM (*Word2Vec+LGBM*) was prepared to compare the performance of the optimal model. The corpus for training the Word2Vec model consisted of 1.14 million words with definitions in the *Urimalsaem*. The hyperparameters of the Word2Vec model were {*window size = 10, vector dimension = 500, epochs = 30*}, and the hyperparameters of LightGBM used the default value of LightGBM Scikit-learn API[2].

Finally, to visualize the performance of the models, 12 words were sampled as a test set from AL_t. The training set consisted of the remaining words except for these words.

Emotion Evaluation on Dictionary

The V-A values of headwords in the Korean dictionary were evaluated. About 440,000 words were selected, excluding words that have no meaning on their own (prepositional particles, prefix, etc.), or groups of names of people, places, or books, from about 1.14 million headwords from *Urimsaim*. The V-A values of each word were evaluated with the model. The model used during this process was trained using the whole data in AL_t without excluding any words for the test set.

Result and Discussion

The test results are shown in Figure 4. The *MSE* of *Word2Vec+LGBM* was significantly higher than the *MSE* of the other two models, particularly in the V dimension. The result of *BERT-optimal* showed improvement in all *MSE* compared to those of *Word2Vec+LGBM*, and MSE_v was less than MSE_a. The result of *ELECTRA-optimal* showed improvement in all indicators, MSE_t, MSE_v, MSE_a, compared those of *BERT-optimal*. In the case of *게시판Bulletin Board*, this was the never seen word to the model, but the V-A values of this word were predicted almost accurately.

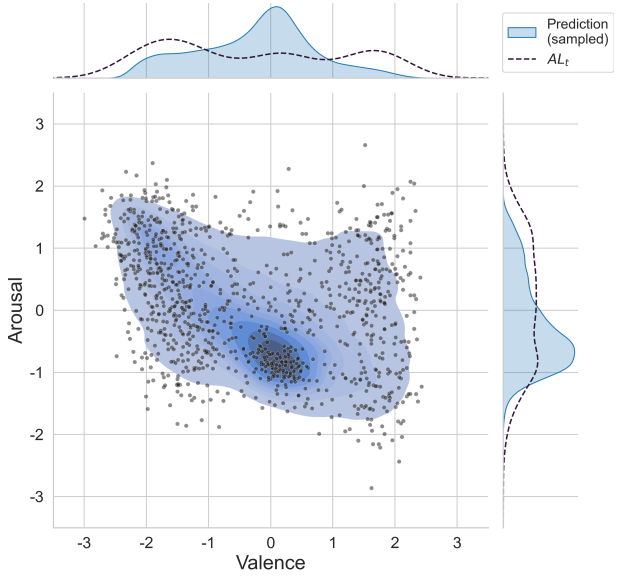

Figure 5: The prediction of V and A values of words in Korean dictionary.

The result of the V-A evaluation for the dictionary is shown in Figure 5. Since all 440,000 words could not be visualized, 10% of all words were sampled and visualized. In visualization, most Korean words appeared to be neutral in terms of the V-A dimension. The neutrally centered distribution reflects that most of the headwords in the dictionary have no emotional meaning. These neutral words appeared around 0 in the V dimension, but -1 in the A dimension. It seems to be because most neutral words in the AL_t were from Hong et al. (2016), and the A values of these words were distributed around -1. In the case of the V dimension, there were more negative words (V<0) than positive words (V>0). It implies that there were aforementioned negative bias also in the Korean vocabulary set.

Study 2

In Study 2, we built a new affective lexicon that excluded ambiguity in previous studies in the evaluation process. We used it to validate the model proposed in Study 1, and then discussed the similarities and differences between human ratings and model predictions by comparing those evaluations.

Building the list of words

To build a new affective lexicon, we filtered the words from 440,000 words selected in Study 1. Filtering criteria was the frequency of the word used in news articles, the meaning of the word, and the predicted V-A value of the word. During this process, words with multiple meanings or homonyms were considered different words. After reviewing by two researchers in this study, 252 words were finally selected.

[2]https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMModel.html#lightgbm.LGBMModel

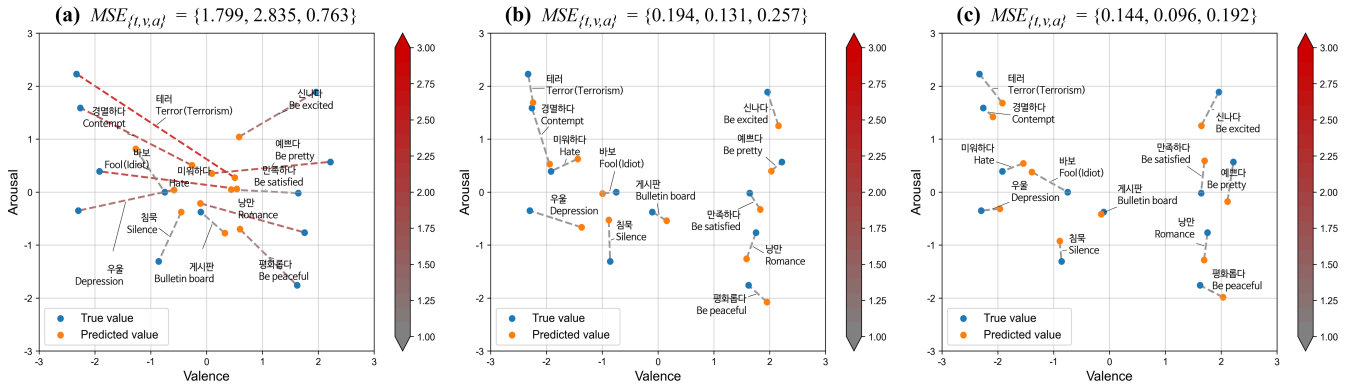

Figure 4: The test results of models. (a) *Word2Vec+LGBM*, (b) *BERT-optimal*, (c) *ELECTRA-optimal*. A dashed line is a distance between true and predicted values. The color of the dashed line is the visualization of the line length. $MSE_{\{t,v,a\}}$ means the mean squared error of $\{V\text{-}A, V$, and $A\}$ dimension.

Evaluating V-A Values of Words With Contextual Meaning

The total word list was split into two lists, and each list have 126 words. The evaluation procedure was conducted via Google Forms using the method in Hong et al. (2016). The informed consent form and explanation on how to evaluate were presented with five pilot examples. After practice, 126 words were presented with a dictionary definition in the form of $\{$*Word: Definition*$\}$, and evaluated. Since words with multiple meanings were previously considered as different words, only one definition was presented with a word. There was a total of 30 participants in the experiment, and each of 15 participants was assigned to two groups of half-divided word list. The participant's gender was 13 males and 17 females, and the ages range from 19 to 36.

Result and Discussion

As a result of evaluating emotion of the words containing contextual meanings, there was no data that had to be excluded in both 252 words and 30 participants. Thus, all words and ratings were used for analysis. The V and A values of a word were determined as the average values of 15 people who evaluated the word. The $\{min, max, mean(std)\}$ of the evaluated words were {-2.87, 2.67, 0.09 (1.44)} in the V dimension, and {-1.87, 2.20, 0.36 (0.88)} in the A dimension. The $mean(std)$ of gender difference was 0.52 (0.42) in the V dimension, and 0.89 (0.73) in the A dimension. The gender difference was larger in the A dimension than in the V dimension. After sorting words by gender differences in descending order, 29 of the 30 words were shown in the A dimension. In the V dimension, there was only one word in which gender difference was significantly large; *페미니스트feminist*.

The human ratings and model predictions of the words are as shown in Figure 6. The error between two group was $MSE_{\{t,v,a\}}$ = {0.71, 0.83, 0.60}. The distribution of human ratings and model predictions seemed similar in the A dimension, but the difference between the two distributions was large in the V dimension. The distribution of model predictions has more negative words than positive words, like the distribution of AL_1 in Figure 1, while the distribution of human ratings was more positively biased than that of the model predictions.

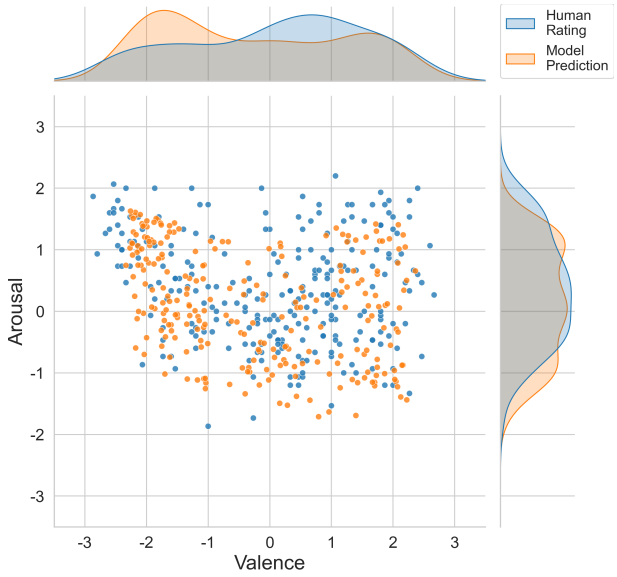

Figure 6: The visualization of human ratings and model predictions

The distribution of V-A values of the words in human ratings was similar to that of previous studies. There were differences in evaluation between genders, and most words with a large difference between genders in the A dimension. Females tended to give higher scores in the A dimension than males in most words. These facts were consistent with the study of Hong et al. (2016). In the V dimension, *페미니스트feminist* was evaluated as a highly negative word to males (V = -2.57) but it was evaluated as a neutral word to females (V = -0.25). On the other hand, *인종주의racism* was evaluated as a negative word to both male (V = -2.00) and female (V = -2.75). These results could be interpreted as a reflec-

Table 3: Examples of words and definitions in which the model showed negative bias in the Valence dimension.

Word: Definition	Human rating	Model prediction
비흡연: 담배를 피우지 **아니**함. **Non**-smoking: **No** smoking.	2.00	-1.11
미혼자: **아직** 결혼하지 **않은** 사람. Single: Someone who is **not** married **yet**.	0.33	-1.71
빈자리: 사람이 앉지 **아니**하여 비어 있는 자리. Empty seat: An empty seat because **no** one sits.	0.27	-1.35

tion of the severe gender conflicts in Korea and the cultural characteristics of Korea consisting of a single ethnic group. This phenomenon corresponds to the claim "emotional experiences by emotion words vary according to culture" mentioned at the Hong et al. (2016).

Meanwhile, there was a large difference between the human ratings and model predictions. The $MSE_{\{t,v,a\}}$ was significantly larger than that found in the model optimization procedure or test result of the models. In particular, MSE_v was larger than our expectation. After reviewing the words with large MSE_v, we figured out the cause of this error. The model negatively predicted the V values of some words containing specific expressions (Table 3).

The result seems to be due to the lack of words in various contexts in the AL_t used to train the model. In many cases, for most negative words, expressions such as 않-, 아니- (*Not-* or *Un-* in English) are included in the definition of words. Also, most of the words in AL_t were words with strong emotional meaning. Therefore, since the variation of the context of data was not enough to train the model and the contextual meanings of the words were (negatively) biased, it was estimated that the model has learned these kinds of (negative) expressions as important features to predict negative words, resulting in this error.

General Discussion

The present study was conducted to expand the Korean affective lexicon with a neural network model using the dictionary definition of words as an input. In Study 1, various structures and hyperparameters were tested to optimize the model. The optimal model was built through 10-fold cross-validation to make up for the small data size. After optimization, we explored the distribution of the emotional meaning of words in the Korean dictionary. Most of the words were emotionally neutral. Values were evenly distributed in the Arousal dimension, but in the Valence dimension, negative words were found more than positive words. It means that the negative bias was also observed in the words we were using. In Study 2, we compared the human ratings with the model predictions to validate the proposed model. In this process, the predicted values on the Valence-Arousal dimension were used to select the words to be evaluated. In order to avoid the limitations of previous studies, the evaluation was conducted in a way that limits the contextual meaning by presenting words as a single dictionary definition. As a result, we found most of the findings regarding human ratings followed the results of previous studies. However, there was a larger-than-expected error between the human ratings and model predictions especially on the Valence dimension. After a deep review of this case, it was assumed that this error was caused by a lack of diversity in the data used to train the model.

The current study has the following significance. First of all, we proposed a method to expand the affective lexicon. Unlike English-speaking countries, where tens of thousands of affective words exist, there are no such large-scale studies in Korean. Thus, the significance of this study is that it has integrated the results of existing research on the Korean affective lexicon and utilized them to expand the affective lexicon without human ratings. Second, we applied the state-of-the-art method to our model. Previous studies used the statistical analysis or word embedding based methods. However, since our model was based on the deep-learning techniques, the emotions of words could be inferred from the dictionary definitions. Lastly, the Valence and Arousal values of all words in the dictionary could be evaluated by simply entering it in the form of *Word: Definition*, even if the words currently not be included in the dictionary. Therefore, our study is meaningful in that the accuracy of emotion predictions and the possibility of expansion of affective lexicon has increased compared to previous studies.

Nevertheless, there are limitations to this study. Although the model was optimized while preventing overfitting, the model made biased predictions in some words including specific expressions due to the size of training data. As a result, there was a larger error than we have expected on the Valence dimension. To overcome this problem, more words in various contexts are needed to train the model. Meanwhile, when testing various structures and hyperparameters combinations during the model optimization procedure, the model was optimized by grid search and trial-and-error. This process took a considerable amount of time and we could have missed a better combination. Therefore, using better optimization techniques, such as Bayesian optimization or AutoML, is needed to optimize the model in future studies.

Acknowledgments

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2021R1A6A3A13039453). We also thank Ji Won Yang for her extremely helpful comments on an earlier draft.

References

Adolph, D., & Alpers, G. W. (2010). Valence and arousal: a comparison of two sets of emotional facial expressions. *The American Journal of Psychology*, *123*(2), 209–219.

Anders, S., Lotze, M., Erb, M., Grodd, W., & Birbaumer, N. (2004). Brain activity underlying emotional valence and arousal: A response-related fmri study. *Human brain mapping*, *23*(4), 200–209.

Bradley, M. M., & Lang, P. J. (1999). *Affective norms for english words (anew): Instruction manual and affective ratings* (Tech. Rep.). Technical report C-1, the center for research in psychophysiology ....

Clark, K., Luong, M.-T., Le, Q. V., & Manning, C. D. (2020). *Electra: Pre-training text encoders as discriminators rather than generators.* arXiv. Retrieved from https://arxiv.org/abs/2003.10555 doi: 10.48550/ARXIV.2003.10555

Clore, G. L., Ortony, A., & Foss, M. A. (1987). The psychological foundations of the affective lexicon. *Journal of personality and social psychology*, *53*(4), 751.

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: covariation with autonomic arousal and affective report. *Biological psychology*, *52*(2), 95–111.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv. Retrieved from https://arxiv.org/abs/1810.04805 doi: 10.48550/ARXIV.1810.04805

Dolcos, F., LaBar, K. S., & Cabeza, R. (2004). Dissociable effects of arousal and valence on prefrontal activity indexing emotional evaluation and subsequent memory: an event-related fmri study. *Neuroimage*, *23*(1), 64–74.

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, *6*(3-4), 169–200.

Hong, Y., Nam, Y. E., & Lee, Y. (2016). Developing korean affect word list and it's application. *Korean Journal of Cognitive Science*, *27*, 377–406.

Kim, B., Lee, E., Kim, H., Park, J., Kang, J., & An, S. (2010). Development of the korean affective word list. *Journal of Korean Neuropsychiatric Association*, *49*(5), 468–479.

Kim, K. (2020). *Pretrained language models for korean.* https://github.com/kiyoungkim1/LMkor. GitHub.

Lazarus, J. (2021). Negativity bias: An evolutionary hypothesis and an empirical programme. *Learning and Motivation*, *75*, 101731.

Mohammad, S. (2018). Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. In *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 174–184).

Monnier, C., & Syssau, A. (2014). Affective norms for french words (fan). *Behavior research methods*, *46*(4), 1128–1137.

National Institute of Korean Language. (2008). *Standard korean language dictionary 표준국어대사전.* Retrieved 2021-12-15, from https://stdict.korean.go.kr/

National Institute of Korean Language. (2016). *Urimalsaem 우리말샘.* Retrieved 2021-12-15, from https://opendict.korean.go.kr/

Ortony, A., Clore, G. L., & Foss, M. A. (1987). The referential structure of the affective lexicon. *Cognitive science*, *11*(3), 341–364.

Palogiannidi, E., Koutsakis, P., Losif, E., & Potamianos, A. (2016). Affective lexicon creation for the greek language.

Park, I. J., & Min, K. H. (2005). Making a list of korean emotion terms and exploring dimensions underlying them. *Korean Journal of Social and Personality Psychology*, *19*, 109–129.

Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, *89*(4), 344–350.

Recchia, G., & Louwerse, M. M. (2015). Reproducing affective norms with lexical co-occurrence statistics: Predicting valence, arousal, and dominance. *Quarterly journal of experimental psychology*, *68*(8), 1584–1598.

Research Institute of Korean Studies. (2009). *Korea university korean dictionary*. 145 Anam-ro, Seongbuk-gu, Seoul 02841, Korea.

Reuderink, B., Mühl, C., & Poel, M. (2013). Valence, arousal and dominance in the eeg during game play. *International journal of autonomous and adaptive communications systems*, *6*(1), 45–62.

Rhee, S. Y., & Ko, I. J. (2013). Measuring a valence and activation dimension of korean emotion terms using in social media. *Science of Emotion and Sensibility*, *16*, 167–176.

Russell, J. A. (1978). Evidence of convergent validity on the dimensions of affect. *Journal of personality and social psychology*, *36*(10), 1152.

Russell, J. A. (1983). Pancultural aspects of the human conceptual organization of emotions. *Journal of personality and social psychology*, *45*(6), 1281.

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of research in Personality*, *11*(3), 273–294.

Sedoc, J., Preoţiuc-Pietro, D., & Ungar, L. (2017). Predicting emotional word ratings using distributional representations and signed clustering. In *Proceedings of the 15th conference of the european chapter of the association for computational linguistics: Volume 2, short papers* (pp. 564–571).

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: the negativity bias in social-emotional development. *Psychological bulletin*, *134*(3), 383.

Vankrunkelsven, H., Verheyen, S., De Deyne, S., & Storms, G. (2015). Predicting lexical norms using a word association corpus. In *Proceedings of the 37th annual conference of the cognitive science society* (pp. 2463–2468).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008).

Võ, M. L., Jacobs, A. M., & Conrad, M. (2006). Cross-validating the berlin affective word list. *Behavior research methods*, *38*(4), 606–609.

Wang, J., Yu, L.-C., Lai, K. R., & Zhang, X. (2016). 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.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological bulletin*, *98*(2), 219.

Yu, L.-C., Lee, L.-H., Hao, S., Wang, J., He, Y., Hu, J., ... Zhang, X. (2016). Building chinese affective resources in valence-arousal dimensions. In *Proceedings of the 2016 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 540–545).