

Research Article

Multilabel Emotion Recognition Technique considering the Characteristics of Unstructured Conversation Data

Myungjin Lim,¹ Moungho Yi,² Pankoo Kim ,¹ and Juhyun Shin ³

¹Department of Computer Engineering, Chosun University, Gwangju 61452, Republic of Korea

²Department of Electronic Engineering, Chosun University, Gwangju 61452, Republic of Korea

³Department of new Industry Convergence, Chosun University, Gwangju 61452, Republic of Korea

Correspondence should be addressed to Juhyun Shin; jhshinkr@chosun.ac.kr

Received 4 February 2022; Revised 5 April 2022; Accepted 25 April 2022; Published 9 May 2022

Academic Editor: Chun-Wei Tsai

Copyright © 2022 Myungjin Lim et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the recent increase in non-face-to-face services due to COVID-19, the number of users communicating through messengers or SNS (social networking service) is increasing. As a large amount of data is generated by users, research on recognizing emotions by analyzing user information or opinions is being actively conducted. Conversation data such as SNS is freely created by users, so there is no set format. Due to these characteristics, it is difficult to analyze using AI (artificial intelligence), which leads to a decrease in the performance of the emotion recognition technique. Therefore, a processing method suitable for the characteristics of unstructured data is required. Among the unstructured data, most emotion recognition in Korean conversation recognizes a single emotion by analyzing emotion keywords or vocabulary. However, since multiple emotions exist complexly in a single sentence, research on multilabel emotion recognition is needed. Therefore, in this paper, the characteristics of unstructured conversation data are considered and processed for more accurate emotion recognition. In addition, we propose a multilabel emotion recognition technique that understands the meaning of dialogue and recognizes inherent and complex emotions. A deep learning model was compared and tested as a method to verify the usefulness of the proposed technique. As a result, performance was improved when it was processed in consideration of the characteristics of unstructured conversation data. Also, when the attention model was used, accuracy showed the best performance with 65.9%. The proposed technique can contribute to improving the accuracy and performance of conversational emotion recognition.

1. Introduction

With the development of the Internet and the increase of non-face-to-face services due to COVID-19, the number of cases of communication through messengers or SNS (social networking service) is increasing. As a large amount of data is generated by users, research to analyze user information and opinions is being actively conducted [1, 2]. Among them, emotion recognition is a research field that recognizes emotions by analyzing personal opinions about social issues or products [3, 4]. However, due to the nature of non-face-to-face communication, misunderstandings may occur unexpectedly as they are interpreted differently between the sender and the receiver. And sometimes the contextual context intended by the sender is not communicated to the

receiver. Transparent emotional exchange is an important factor for dialogue without misunderstanding between the parties involved.

Text data is the most developed emotional expression means that can more precisely express the various emotions that humans experience. Recognizing the emotions conveyed by textual expressions is essential for machines to read human emotions. In order for AI (artificial intelligence) to accurately grasp the emotional responses contained in text expressions, it is essential to have a rich and sophisticated database of human emotional expression languages. And since the emotions that humans experience in social situations are much more diverse than the basic emotions, it is necessary to classify as many emotions as possible [5]. Recognizing the user's emotions from text data allows us to

analyze the user's intentions in various ways [6]. It also helps to understand the user's point of view.

There is no set format for conversations used on the Internet. Among these unstructured conversational data, if we look at the characteristics of Korean, many new words and profanities are generated and used frequently [7]. In addition, the use of abbreviations that freely transformed Korean is increasing. The use of such unstructured data sometimes causes communication problems because users other than the main user group have difficulties in understanding words and meanings. In addition, since it does not have a dictionary meaning, performance degradation and limitations of the emotion recognition algorithm occur [8]. Therefore, it is necessary to identify and process the characteristics of unstructured Korean, which includes words such as neologisms, profanities, and abbreviations. Another characteristic of unstructured Korean is that it is difficult to analyze because it is different from English. Therefore, it is important to understand the order of words and the meaning of the entire sentence. If only the word is used as a feature, data sparseness may occur depending on the characteristics of Korean, where morphological transformations occur in various ways. Therefore, a better understanding of the characteristics of unstructured Korean conversation data and a specialized processing method are needed.

Text emotion classification studies mainly recognize emotions through vocabulary. Alternatively, emotions in word or sentence units are analyzed using an emotion dictionary or a large number of uni-grams. However, in this classification, it is important to balance the quality of training data, the amount of training data, and the data [9]. In addition, as a study of emotion classification using models, attempts to improve emotion classification performance using CNN models, LSTM models, and CNN-LSTMs continued [10–12]. However, research that improves quality by improving data rather than improving the model is more effective in terms of performance [13].

Multilabel emotion recognition in conversations is an important task in providing the ability to understand the user's various emotions. As a related study, using MEISD (Multimodal EmotionLines Dataset), it was proved that the results of multilabel emotion recognition of text are more accurate than other inputs based on DialogueRNN [14]. In addition, using SemEval-2018, multiple emotions of sentences were recognized based on AttnConvnet [15]. And we applied a content-based method to tweets to classify multilabel emotions [16]. Attempts to recognize multiple emotions in various ways are continuing, but most of them were studies using English data sets. Most of the studies using the Korean conversation data set have a limitation in that they are classified as a single emotion. Therefore, considering the characteristics of the Korean language, it is necessary to understand the meaning of dialogue and recognize the underlying emotions. And we need a way to recognize the various emotions included in the sentence.

Therefore, in this paper, the following technique is proposed to solve the processing problem suitable for unstructured Korean conversation data and the single emotion recognition problem. It is a data processing method that considers the characteristics of unstructured conversation

data for more accurate emotion recognition and a multilabel emotion recognition method applying it. It contributes to improving the accuracy and performance of emotion recognition by processing unstructured data and recognizing various emotions inherent in sentences. The structure of this paper is as follows. The related study describes emotion recognition of unstructured texts and emotion analysis in Korean, and the proposed method describes a multilabel emotion recognition technique that considers the characteristics of unstructured conversation data. In the experimental results, the proposed technique is tested and evaluated, and finally, conclusions and future research are described and finalized.

2. Related Work

In text emotion recognition, when a machine recognizes human emotions, it classifies emotions on current input data. In addition, beyond this level, it is necessary to enable more accurate emotional recognition intelligently according to past memories, emotional subjects, personality, or inclinations [17]. In the past, most emotions were judged by extracting emotion keywords. Since this method loses various syntactic or semantic information contained in natural language sentences, there is a limit to recognizing complex human thoughts such as emotions.

Unstructured text emotion classification studies mainly try to recognize emotions through vocabulary. In the study of [18], an emotion word dictionary and an emotion emoticon dictionary corresponding to Ekman's six emotion categories were constructed. Based on this, we tried to recognize emotions in the blog text by using the SVM classifier. In the study of [19], the data from Twitter, a representative social network service, were analyzed as the subject of analysis. Instead of constructing and utilizing an emotion dictionary for emotion classification, a large number of uni-grams were used. However, in the classification of the supervised learning method using dictionary or vocabulary features, the problem of insufficient data is the most problematic. Various emotion classes and vocabulary used for each class appear in various ways [20]. Therefore, it is very difficult to construct sufficient learning data to enable learning at an appropriate level using these vocabularies. Therefore, it is important to balance the quality of training data, the amount of training data, and the data.

In the study of [21], as a study on the analysis of Korean emotions, emotions were analyzed using a dictionary of emotional words centered on Korean words like English. In this case, there is a disadvantage that one word can fall into the problem of ambiguity in which it is interpreted as having more than one meaning. Considering this, in the study of [22], emotion was learned by focusing on the order of words without using part-of-speech tagging or emotion word dictionary. We present an algorithm with high accuracy in Korean that compares and analyzes the pattern of word order and input sentences. However, in the case of having a clear emotional word rather than an ambiguous expression, the accuracy was rather low. Therefore, not only the order of words but also the use of the emotional word dictionary is important in the analysis of Korean emotions.

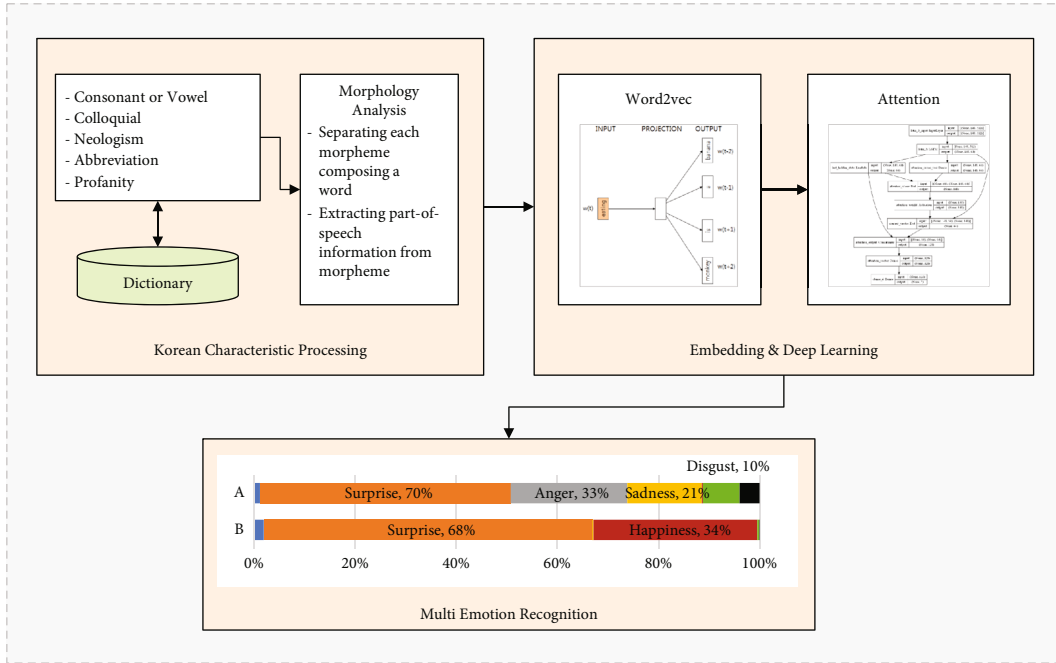


FIGURE 1: System architecture.

TABLE 1: Rules for processing unstructured words.

| Division | Original data → processing data |
|--------------------|---|
| Consonant or vowel | k → keu, h → heu, d → da, u → ngu, yu → ngyu |
| Colloquial | ~dang → ~ da, ~ngyos → ~ ngyo, ~neong → ~ neong |
| Neologism | romgok → tear, TMI → too much information |
| Abbreviation | masang → pain in the heart |
| Profanity | kkmsi → the technique is vicious and cunning |

```

Input: OriginalDataList
Output: ProcessDataList
Definition: Dictionary – standard word definition for unstructured words
                [consonant or vowel, colloquial, abbreviation, neologism, profanity]
1. READ OriginalDataList;
2. for each sentence in OriginalDataList do
3.   for each word in sentence do
4.     if word in Dictionary then
5.       word ← standardword;
6.     end if
7.   end for
8. end for
9. ProcessDataList ← OriginalDataList
10. return ProcessDataList;
    
```

ALGORITHM 1: Korean characteristic processing.

And in order to solve the ambiguity problem, the entire sentence should be used for emotion analysis.

As a study of emotion classification using deep learning, in the study of [10], the CNN model showed good performance in sentence and document classification with simple convolution and pooling. In addition, in the study of [11],

the LSTM model was trained considering the order of continuous input data. In terms of generation, it shows good performance not only in machine translation but also in various problems. In the study of [12], context was reflected within the classification model. In addition, the CNN-LSTM model was used to classify speech emotions in

```

Input: d:dataset
Output: Matrix  $W_{(512,113)}$  of one-hot vectors for each possible byte value (0-255)
1. let f be a list of tuples (byte_value, frequency);
2. for i :=0 to 511 do
3.   freq  $\leftarrow$  0;
4.   for each item j in d do
5.     freq  $\leftarrow$  freq + frequencyOfOccurrence(i,j);
6.   end for
7.   append (i,freq) tuple to f;
8. end for
9. f  $\leftarrow$  sort f based on frequencies;
10. W  $\leftarrow$  word2vec(f,113);
11. return W;

```

ALGORITHM 2: Word2vec embedding.

```

Input: d:dataset, l:dataset true lables, W:word2vec matrix
Output: ME of Attention trained model on test dataset
Definition: ME – multiemotion
1. let f be the featureset 3d matrix;
2. for i in dataset do
3.   let  $f_i$  be the featureset matrix of sample i;
4.   for j in i do
5.      $v_j \leftarrow$  vectorize(j,w)
6.     append  $v_j$  to  $f_i$ 
7.     append  $f_i$  to f
8.   end for
9. end for
10.  $f_{train}, f_{test}, l_{train}, l_{test} \leftarrow$  split feature set and lables into train subset and test subset
11. M  $\leftarrow$  Attention ( $f_{train}, l_{train}$ )
12. ME  $\leftarrow$  evaluate ( $f_{test}, l_{test}, M$ )
13. return ME

```

ALGORITHM 3: Attention model.

TABLE 2: Number of data by emotion class.

| Data set | Neutral | Surprise | Anger | Sadness | Happiness | Disgust | Fear | Total |
|------------|---------|----------|--------|---------|-----------|---------|--------|--------|
| Single | 4,830 | 5,898 | 5,665 | 5,267 | 6,037 | 5,429 | 5,468 | 38,594 |
| Continuous | 43,786 | 4,866 | 3,628 | 1,972 | 1,030 | 220 | 98 | 55,600 |
| Merge | 48,616 | 10,764 | 9,293 | 7,239 | 7,067 | 5,649 | 5,566 | 94,194 |
| (%) | (51.6%) | (11.4%) | (9.9%) | (7.7%) | (7.5%) | (6.0%) | (5.9%) | (100%) |

conversations to enable automatic learning by extracting a large amount of vocabulary. All emotion classification performance was improved, but there was a limitation in that they were classified as a single emotion.

As a study on multilabel emotion recognition, [14] used MEISD (Multimodal EmotionLines Dataset) to recognize multilabel emotion based on DialogueRNN. It was proved that the results of multilabel emotion recognition of text are more accurate than that of audio and video. In the study of [15], SemEval-2018 was used to recognize multiple emotions of a sentence based on AttnConvnet, which combines attention and convolution. The two-step procedure used by humans to analyze a sentence was first to understand the meaning of the sentence and to

classify the emotions. In the study of [16], multilabel sentiment was classified by applying a content-based method (word and character n-gram) to tweets. We showed that the content-based word unigram outperformed other methods. However, most of the studies were conducted using English data. Most of the studies using the Korean conversation data set were classified as a single emotion. Therefore, it is necessary to understand the meaning of conversation in consideration of the characteristics of the Korean language and to recognize the various emotions included in the sentence by recognizing the emotions embedded in it.

Most of the AI research that automatically recognizes and extracts emotions from text has been model-centric.

TABLE 3: Comparison of emotion recognition performance by data set.

| Data set | Original data | Process data | Performance comparison |
|------------|---------------|--------------|------------------------|
| Single | 43.2% | 56.4% | 13.2% |
| Continuous | 40.7% | 51.6% | 10.9% |
| Merge | 47.1% | 65.9% | 18.8% |

But beyond research, the most important thing in AI development in practice was data, not models. How the data is preprocessed, how it is collected, what the size is, how good the quality is, and how the training/evaluation set is divided greatly affect the development of AI systems. In reality, it is the improvement of the data, not the improvement of the code, that increases the performance of the AI system [13]. Therefore, the method to improve the model is important, but the study that improves the quality by improving the data is more effective in terms of performance. In this paper, as a method to improve data, it is processed in consideration of the data imbalance resolution and the characteristics of unstructured conversation data. And, as a way to improve the model, we propose a multilabel emotion recognition technique in sentences using LSTM with attention.

3. Proposed Method

3.1. System Architecture. This section describes the data processing method and multilabel emotion recognition method considering the characteristics of the proposed unstructured conversation data. Figure 1 is a system configuration diagram of the proposed method. As a data set processing process, the characteristics of unstructured conversation data are considered and processed. It converts ellipsis, colloquial words, neologisms, abbreviations, and slang words into standard words and analyzes morphemes. The word vectorization is carried out by learning the morpheme-analyzed sentence using the Skip-gram method of the Word2vec model. It trains a deep learning model with learned word vectors and emotion classes and recognizes multiple emotions in conversational sentences through prediction.

3.2. Feature Processing of Unstructured Conversation Data. This section describes a data processing method considering the characteristics of unstructured conversation data. The unstructured Korean data set has language destructive features such as “consonant or vowel,” “colloquial,” “neologisms,” “abbreviations,” and “profanity.” As a result, the morpheme analyzer that separates sentences by part-of-speech does not work properly [19]. Therefore, it is necessary to process the data considering the characteristics of unstructured data. In this paper, “consonant or vowel,” “colloquial,” “neologisms,” “abbreviations,” and “profanity” are defined as unstructured words. Table 1 shows an example of a processing rule for an unstructured word.

As shown in Table 1, “consonant or vowel” using only consonant or vowel such as “k” and “u” is converted to

“keu” and “ngu,” which are forms of “consonant + vowel” according to the characteristics of Korean. It converts “colloquial” words that are frequently used online, such as “~neong,” into standard words such as “~neo.” A “neologism” is converted to a standard language by referring to the dictionary of neologism [23] provided by the Naver Open Dictionary PRO and the neologism list [24] provided by Wikipedia. The “abbreviation” is converted to standard language by referring to the list of abbreviations on Namu Wiki [25]. “Profanity” is converted to standard language by referring to the Standard Korean Dictionary of the National Institute of the Korean Language [26]. According to these processing rules, a standard word suitable for an unstructured word is matched and stored in the unstructured word dictionary. Next, morphological analysis is performed. There were many sentences in which spaces were ignored in the data set. For this reason, the Okt (Open Korean Text) morpheme analyzer was used, which is good for analyzing sentences without spaces in Korean data.

Algorithm 1 shows the unstructured Korean processing rule algorithm used in this paper. Extract a sentence with an original data set (OriginalDataList) as input, and extract a word from the sentence. Next, it is checked whether the word is included in the dictionary (dictionary: defines standard words for unstructured words). If it exists in the dictionary, it is converted to a defined standard word. The original data set converted to the standard language is stored in the process data set (ProcessDataList).

3.3. Word2vec Embedding. This section describes the process of learning using the Word2vec model on data processed considering the characteristics of unstructured Korean. As unstructured Korean is highly dependent on words, it is important to understand the order of words and the meaning of the entire sentence. Therefore, as the embedding method, the Word2vec model, which can broadly grasp the meaning by considering the context before and after, was used. In order to understand the meaning according to the order of the words, the entire sentence of the data is trained with the Word2vec model to vectorize the words.

Algorithm 2 shows the algorithm of the Word2vec embedding model used in this paper. The learning method used in the Word2vec model used Skip-Gram, which predicts the context-word from the center-word. As hyperparameters, set the Learning Rate to 0.05, Dimension (Vector space) to 512, Window Size to 2, and Min Count to 5. Through embedding, it is possible to extract semantically similar words from data. And by using vector values as input values, deep learning becomes possible.

3.4. Multilabel Sentiment Recognition. In this section, we describe the process of learning vector and sentence emotion classes learned with the Word2vec model for multilabel emotion recognition using deep learning. Attention is used as a deep learning model to predict the complex emotions of a given sentence. Attention is a model that re-references the encoder’s input sequence every time the decoder predicts the output word. At this time, an attention value is additionally required because the input sequence is viewed more

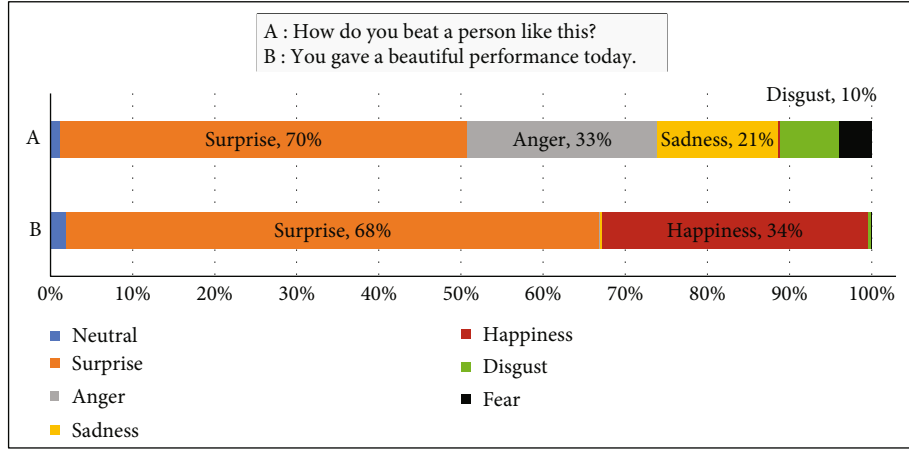


FIGURE 2: Results of multiemotion recognition.

TABLE 4: Comparative evaluation table for each emotion.

| Emotion | LSTM | | | CNN-LSTM | | | Attention | | |
|-----------|-----------|--------|----------|-----------|--------|----------|-----------|--------|----------|
| | Precision | Recall | F1-score | Precision | Recall | F1-score | Precision | Recall | F1-score |
| Neutral | 0.387 | 0.387 | 0.387 | 0.387 | 0.444 | 0.410 | 0.459 | 0.470 | 0.459 |
| Surprise | 0.442 | 0.475 | 0.453 | 0.479 | 0.559 | 0.513 | 0.565 | 0.588 | 0.577 |
| Anger | 0.475 | 0.486 | 0.486 | 0.536 | 0.502 | 0.513 | 0.600 | 0.624 | 0.612 |
| Sadness | 0.464 | 0.585 | 0.519 | 0.559 | 0.582 | 0.571 | 0.636 | 0.671 | 0.659 |
| Happiness | 0.695 | 0.706 | 0.695 | 0.755 | 0.732 | 0.743 | 0.824 | 0.801 | 0.813 |
| Disgust | 0.618 | 0.486 | 0.541 | 0.686 | 0.536 | 0.605 | 0.765 | 0.683 | 0.718 |
| Fear | 0.640 | 0.629 | 0.640 | 0.640 | 0.686 | 0.663 | 0.765 | 0.754 | 0.754 |
| Average | 0.532 | 0.536 | 0.532 | 0.577 | 0.577 | 0.574 | 0.659 | 0.656 | 0.656 |
| Accuracy | | 0.533 | | | 0.577 | | | 0.659 | |

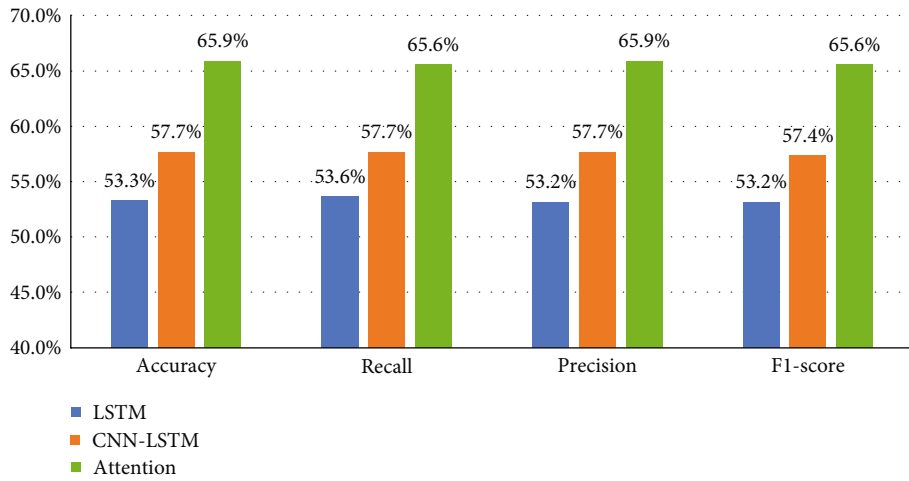


FIGURE 3: Comparison graph of the evaluation methods.

closely as the input word related to the predicted word. Equations (1) to (3) represent the equations for obtaining the attention value used in this paper:

$$\text{score}(s_t, h_i) = s_t^T h_i, \quad (1)$$

$$\alpha^t = \text{softmax}(e^t), \quad (2)$$

$$a_t = \sum_{i=1}^N \alpha_i^t h_i. \quad (3)$$

Equation (1) shows a method to obtain the attention score of the hidden state s_t at time step t of the decoder and the hidden state h_i at time step i of the encoder. Equation (2) shows a method of obtaining the attention distribution, a probability distribution in which the sum of all values becomes 1 by applying the softmax function to the attention score obtained in Equation (1). Attention distribution obtains a vector in the form of [0.1, 0.4, 0.1, 0.3], and each value is called attention weight. Equation (3) shows the method of obtaining the attention value through the attention weight obtained in Equation (2) and each hidden state. Finally, the attention value a_t is connected to the hidden state at time t of the decoder.

Algorithm 3 shows the algorithm of the attention model used in this paper. The designed model is composed of attention to solve the vanishing gradient problem, which is a disadvantage of LSTM. The input data is the number of collected sentences and 113, which is the maximum number of words per sentence. It is a 512-dimensional embedded Word2vec model that learns words and consists of 3 dimensions (number of sentences, word length, and embedding vector). Attention is located in the output part of LSTM so that it can be predicted well even if the length of the sentence is long, and 128 units are configured. In the output layer, to predict multiple emotions, the output of Dence was set according to the number of emotions, and softmax was used as the activation function.

Using the attention model, you can understand the meaning of a sentence, so you can better understand the dependencies between words. Through the semantic information extraction, it becomes possible to recognize the emotions inherent in sentences and to recognize complex emotions [27]. Even with data classified as a single emotion, using the proposed model, multiple emotions in a sentence can be recognized. Therefore, if the multilabel emotion recognition method is used, it is possible to understand the meaning of the conversation and recognize various emotions inherent in it so that more accurate emotion recognition and performance improvement are possible.

4. Experimental Results

In this chapter, we experiment with the data processing method and the multilabel emotion recognition method considering the characteristics of the proposed unstructured conversation data. In order to prove the usefulness of the data set and multilabel emotion recognition technique used in the experiment, the experimental results and performance evaluation are described.

4.1. Data Set. This section describes the data set used to test and evaluate the performance of the proposed technique. The data used were the “single conversation data set including Korean emotion information” and the “continuous conversation data set” provided by AI-HUB [28]. The two data sets were constructed by labelling the sentences selected by the web crawling on SNS, comments, and conversations, which are unstructured data, as emotion classes. The single conversation data set is divided into “sentence” and “emo-

tion” and consists of a total of 38,594 sentences. Each sentence is classified into seven emotion classes: fear, surprise, anger, sadness, neutral, happiness, and disgust. The continuous dialogue data set is divided into “dialog,” “utterance,” and “emotion.” It consists of a total of 55,600 sentences and is classified into 7 emotion classes. Table 2 shows the number of data for each emotion class.

As shown in Table 2, it can be seen that the data imbalance of the continuous conversation data set is more severe than that of the single conversation data set. In this paper, two data sets are merged and used to construct training data sufficient to learn emotion recognition deep learning. The merged data set (merge) consists of data with an imbalance in the number of data for each emotion class. The most emotions are the neutral class, with 48,616, accounting for 51.6%. The lowest emotion is the fear class, with 5,566, accounting for 5.9%. In this paper, to solve the data imbalance problem, the number of data per emotion class is balanced with the smallest number of 5,500. A total of 38,500 data, 5,500 of each of the seven emotions, were used in the experiment.

4.2. Performance Evaluation. In this section, we describe the performance evaluation results of the data processing method and the multilabel emotion recognition method considering the characteristics of the unstructured conversation data proposed in this paper. For the comparison with multiple data sets, a single conversation data set, a continuous conversation data set, and a merged data set (merge) were used as the original data. Then, unstructured words were converted into standard words, morphologically analyzed, and used as processing data. Next, the words were vectorized by learning the sentences with the Word2vec model. Multiple emotions were recognized through deep learning with the generated vector and emotion class. Based on 5,500 per emotion class, the emotion class that is less than 5,500 was used as it is. In the train, the test data were used in a ratio of 8:2. For each data set, the results of learning with the original data set and the results of learning with the processed data set were compared. Table 3 shows the results of comparing the emotion recognition performance of the original data and the processed data of the three data sets.

As shown in Table 3, it can be seen that the performance of the model trained with the processed data is more improved than the original data for all three data sets. And it can be seen that the performance of the data set that merges the two data sets is the best. In addition, it can be seen that the performance of the single conversation data set is better because the data imbalance of the continuous dialogue data set is more severe. Therefore, for more accurate emotion recognition and performance improvement, a method for balancing data and a processing method considering the characteristics of unstructured conversation data are needed.

We also analyze the results of multilabel emotion recognition. Emotions are often expressed together with other emotions rather than being expressed alone. And most emotions can be classified into positive and negative. “Happiness” is classified as positive and “anger, sadness, disgust, and fear” as

negative. However, “neutral” is an emotion that cannot be classified, and “surprise” is an emotion that is ambiguous to analyze as positive or negative. There are “positive surprise,” and there are “negative surprise.” Figure 2 shows the multi label emotion recognition results of two sentences classified as “surprise.” In the data set, two sentences were classified with the same emotion of “surprise,” but when analyzed semantically, sentence A (How do you beat a person like this?) is negative, and sentence B (You gave a beautiful performance today) is positive. Contains the meaning of as a result of multilabel emotion recognition of two sentences, sentence A contains negative multiple emotions of surprise 70%, anger 33%, sadness 21%, and disgust 10%. On the other hand, it can be seen that sentence B contains multiple positive emotions of surprise 68% and happiness 34%. Also, if the multilabel emotion recognition is applied to sentences classified as “neutral,” even a small amount of emotion can be recognized, making emotion classification possible. Therefore, if the multilabel emotion recognition method is used, it is possible to understand the meaning of the conversation and recognize the inherent emotions and complex emotions so that more accurate emotion recognition and performance improvement are possible.

4.3. Comparative Evaluation. In this section, we describe the results of comparative experiments after learning three deep learning models using the processed data of the data set that merged the two data sets. In the total data, training data and validation data were composed in a ratio of 8:2. Learning through training data and evaluation through validation data were conducted. The deep learning model used LSTM, CNN-LSTM, and attention, which are widely used in text emotion classification studies. As for the evaluation method, accuracy, precision, recall, and F1-score were used focusing on representative emotions. Table 4 shows the comparison evaluation table for emotion recognition, and Figure 3 shows the comparison graph of the four evaluation methods. Among the proposed three deep learning models, the attention model showed the best performance with 65.9%. In the comparative evaluation by emotion class, the classification performance of the “happiness” emotion class was the best than other emotions. This is interpreted because when the seven emotions are classified as positive and negative, the “happiness” class is classified as positive, and other emotions are classified as negative. In addition, the “surprise” class included both positive and negative emotions, so the performance was classified as lower than other emotions. The “neutral” class with ambiguous classification had the lowest performance. Therefore, the data is processed by solving the data imbalance and considering the characteristics of the unstructured conversation data. Then, by classifying the sentences into multilabel emotion classes and recognizing multiple emotions, it is possible to understand the meaning of conversations and to recognize more accurate emotions embedded in the sentences and improve performance.

5. Conclusions

In this paper, for more accurate emotion recognition, a multilabel emotion recognition technique that considers the

characteristics of unstructured conversation data is proposed. The data sets “singlec information” and “continuous conversation data set” provided by the AI-HUB were used. The number of data for each emotion class was composed of unbalanced data, so the number of data was balanced to 5,500. In a data processing process that considers the characteristics of unstructured Korean, unstructured words such as ellipsis, colloquial words, neologisms, abbreviations, and profane words were converted into standard words. After analyzing the morphemes using Okt, the entire sentence of the data set was vectorized by learning it with the Word2vec model. Multiple emotions were recognized by using a deep learning model with vectors and emotion classes learned with the Word2vec model. The data set is classified as a single emotion per sentence. However, using the multilabel emotion recognition method, it is possible to recognize various emotions inherent in sentences, thus showing that more accurate emotion recognition and performance improvement are possible. As a result of performance evaluation of the proposed technique, it was shown that there is a difference in emotion recognition performance depending on whether unstructured data is included and the imbalance of data. And it was seen that the accuracy of the deep learning model trained with processed data was improved by 18.8% compared to the original data. In addition, as a result of comparing accuracy, precision, recall, and F1-score of three deep learning models, attention was 65.9% higher than CNN and CNN-LSTM, showing the highest accuracy. Therefore, the proposed technique resolves the data imbalance and recognizes multiple emotions in a sentence by applying a data processing method that considers the characteristics of unstructured conversation data. As a result, it was proved that it is possible to improve the emotion recognition performance by being able to see the change of emotion embedded in the sentence. By using the multilabel emotion recognition method that considers the characteristics of unstructured conversation data, it is possible to understand the meaning of a conversation and to recognize detailed emotions other than the representative emotions in a sentence so that more accurate emotion recognition is possible. Through these subtle emotional changes, we can recognize the flow of emotions in conversation. This study can contribute to improving the accuracy and performance of conversational emotion recognition. In addition, AI that accurately recognizes emotions can be applied to robots that interact directly with humans and can be used in various fields such as counseling therapy, emotional engineering, emotional marketing, and emotional education. For future research, we plan to build a deep learning model that recognizes multiple emotions in a continuous conversation by specifically reflecting multilabel emotions and build a system to predict continuous emotions.

Data Availability

The data used to support the findings of this study are available at <https://aihub.or.kr/opendata/keti-data/recognition-laguage/KETI-02-009> and <https://aihub.or.kr/opendata/keti-data/recognition-laguage/KETI-02-010>.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (Nos. 2019R1F1A1057325 and NRF-2020R1A2C2007091).

References

- [1] G.-M. Yoon, "Performance improvement of movie recommendation system using genetic algorithm and adjusting artificial neural network parameters," *The Journal of KING Computing*, vol. 10, no. 5, pp. 56–64, 2014.
- [2] J. Seo and J. Park, "Data filtering and redistribution for improving performance of collaborative filtering," *The Journal of KING Computing*, vol. 17, no. 4, pp. 13–22, 2021.
- [3] G. Son, W. Lee, K. Han, and S. Kyeong, "The study of feature vector generation and emotion recognition using EEG signals," *The Journal of KING Computing*, vol. 16, no. 2, pp. 72–79, 2020.
- [4] M. J. Lim and J. H. Shin, "Continuous emotion recognition method applying emotion dimension," in *2021 Spring Conference of KISM*, vol. 10no. 1, pp. 173–175, 2021.
- [5] P. Ji-Eun, S. Sun-Ju, and S. Jin-Hun, "Classification and intensity assessment of Korean emotion expressing idioms for human emotion recognition," *Journal of the Ergonomics Society of Korea*, vol. 31, no. 5, pp. 617–627, 2012.
- [6] S. Kang, H. Park, and J. Seo, "Emotion classification of user's utterance for a dialogue system," *Korean Journal of Cognitive Science*, vol. 21, no. 4, pp. 459–480, 2010.
- [7] S. J. Park, *Sensitivity Identification Method for New Words of Social Media Using the naïve Bayes Classification*, vol. 8, Chosun University, 2018.
- [8] M. H. Yi, M. J. Lim, H. Ko, and J. H. Shin, "Method of profanity detection using word embedding and LSTM," *Mobile Information Systems*, vol. 2021, 9 pages, 2021.
- [9] M. Purver and S. Battersby, "Experimenting with distant supervision for emotion classification," in *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 482–491, 2012.
- [10] Y. Kim, "Convolutional neural networks for sentence classification," in *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1746–1751, 2014.
- [11] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [12] D. Shin, *Emotion Classification in Dialogue Utterances Using CNN-LSTM Complex Model*, vol. 2, Korea University, 2017.
- [13] *From model-centric to data-centric AI development*, 2021, November 2021, <https://jiho-ml.com/weekly-nlp-35/>.
- [14] M. Firdaus, H. Chauhan, A. Ekbal, and P. Bhattacharyya, "MEISD: a multimodal multi-label emotion, intensity and sentiment dialogue dataset for emotion recognition and sentiment analysis in conversations," in *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 4441–4453, 2020.
- [15] K. I. M. Yanghoon, L. E. E. Hwanhee, and J. U. N. G. Kyomin, "AttnConvnet at SemEval-2018 task 1: attention-based convolutional neural networks for multi-label emotion classification," 2018, <https://arxiv.org/abs/1804.00831>.
- [16] I. Ameer, N. Ashraf, G. Sidorov, and H. Gómez Adorno, "Multi-label emotion classification using content-based features in twitter," *Computación y Sistemas*, vol. 24, no. 3, pp. 1159–1164, 2020.
- [17] X. Zeng, Q. Chen, S. Chen, and J. Zuo, "Emotion label enhancement via emotion wheel and lexicon," *Mathematical Problems in Engineering*, vol. 2021, 11 pages, 2021.
- [18] S. Aman and S. Szpakowicz, "Identifying expressions of emotion in text," in *International Conference on Text, Speech and Dialogue. Lecture Notes in Computer Science*, vol. 4629, pp. 196–205, 2007.
- [19] K. Lee, J. Baik, and S. Lee, "Estimating a pleasure-displeasure index of word based on word similarity in SNS," *Journal of Computing Science and Engineering*, vol. 20, no. 3, pp. 159–164, 2014.
- [20] J. Kim, T. Hong, and P. Kim, "Replacing out-of-vocabulary words with an appropriate synonym based on Word2Vec," *Mobile Information Systems*, vol. 2021, 7 pages, 2021.
- [21] S.-Y. Kim and H. Yu, "Korean multiple sensibility analysis technique of SNS unstructured data," in *2018 Summer Conference of KACE*, vol. 22no. 2, pp. 147–149, 2018.
- [22] M. Kang, *Sentiment analysis in text mining using ensemble-HMM*, Hanyang University, 2016.
- [23] "Open dictionary pro," *Naver open dictionary PRO*, 2021, September 2021, <https://open-pro.dict.naver.com/>.
- [24] "List of New Internet Terms in Korea," *Wikipedia*, 2021, September 2021, https://ko.wikipedia.org/wiki/%EB%8C%80%ED%95%9C%EB%AF%BC%EA%B5%AD%EC%9D%98_%EC%9D%B8%ED%84%B0%EB%84%B7_%EC%8B%A0%EC%A1%B0%EC%96%B4_%EB%AA%A9%EB%A1%9D.
- [25] "Abbreviation," *namu wiki*, 2021, November 2021, <https://namu.wiki/w/%EC%A4%84%EC%9E%84%EB%A7%90?from=%EC%A4%80%EB%A7%90>.
- [26] "Standard Korean dictionary," *National Institute of Korean Language*, 2021, October 2021, <https://stdict.korean.go.kr/main/main.do>.
- [27] M. Lim, M. Yi, and J. Shin, "Performance improvement method of deep learning-based emotion recognition," *The Journal of KING Computing*, vol. 17, no. 5, pp. 88–95, 2021.
- [28] "Cognitive Technology-Language Intelligence," *AI Hub*, 2021, February 2021, https://aihub.or.kr/kefi_data_board/language_intelligence.