Word-level Emotion Embedding based on Semi-Supervised Learning for Emotional Classification in Dialogue

1st Young-Jun Lee
School of Computing
KAIST
Daejeon, Republic of Korea
yj2961@kaist.ac.kr

2nd Chan-Yong Park
School of Computing
KAIST
Daejeon, Republic of Korea
ptparty@kaist.ac.kr

3rd Ho-Jin Choi
School of Computing
KAIST
Daejeon, Republic of Korea
hojinc@kaist.ac.kr

***Abstract*—Emotion classification has been remarkable studies in recent years. However, most of works do not consider the context information such as a flow of emotions. In this paper, we propose the emotion classification in dialogue based on the semi-supervised word-level emotion embedding. For the word-level emotion embedding, we use the NRC Emotion Lexicon which is a list of English words and their associations with eight basic emotions. By adding word-level emotion vectors, we obtain an utterance-level emotion vector. We train a single layer LSTM-based classification network in dialogue. Also, we will evaluate our model on the EmotionLines which is dataset with emotions labeling on all utterances in each dialogue. The experiment plan is described in this paper.**

***Index Terms*—emotion classification, word-level emotion embedding, semi-supervised learning, utterance-level emotion representation, single layer LSTM-based classification network**

I. Introduction

As a necessary part of human intelligence, emotional intelligence is defined as the ability to perceive, integrate, understand, and regulate emotions [1]. There have been many studies to make a machine understand human emotions in natural language processing and computer vision. In computer vision, there have been lots of studies based on deep learning models, such as CNN, RNN [2]–[4]. However, in natural language processing, it is difficult to make a machine understand human emotions in text. Because, recognizing human emotions only in sentences is hard to a machine. To do this, a machine should be possible to classify human emotions from utterances and generate a dialogue by using these classified emotions. In this paper, we do not cover the latter. Thus, we will focus on classifying human emotions in dialogue.

The task of emotion classification figures out the emotion of texts whether it is included in Ekman's 6 basic emotions: anger, disgust, fear, happiness, sadness, and surprise [5]. There have been many studies to analyze these emotions in sentences [6], [7]. However, the previous works have problems. The previous works do not consider the context information, such as a flow of emotions, in dialogue.

In this paper, we proposed the emotion classification in dialogue based on the semi-supervised word-level emotion embedding. Before classifying the emotions in dialogue, we train a modified version of skip-gram model to obtain word-level emotion vectors. These vectors were trained by the semi-supervised learning. For labeling word's emotions, we took the NRC Emotion Lexicon which have lists of english words with eight basic emotions and two sentiments [8]. Through the semi-supervised learning, words which are not labeled in the NRC Emotion Lexicon can be represented as emotions in vector space. However, we will only consider seven basic emotions (Ekman's six basic emotions plus the neutral) from this emotion lexicon. To obtain an emotion of utterance, we will add these vectors in a same utterance. After that, we trian a single layer LSTM-based classification network in dialogue. We will evaluate our task on a benchmark which have labeled one of eight emotions (Ekmans's six basic emotions plue the neutral and the non-neutral) [9].

In summary, this paper makes the following contributions:

- It proposes an emotion classification model in dialogue. It has three mechanisms: a word-level emotion embedding, an utterance-level emotion representation, and an emotion classification in dialogue.
- It proposes the word-level emotion embedding model based on the semi-supervised learning. Through this embedding, words can be labeled with emotions.

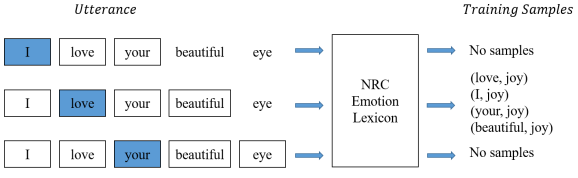

Fig. 1. Sampling data for training.

II. Background

In this section, we briefly describe the previous works used in this paper.

978-1-5386-7789-6/19/$31.00 ©2019 IEEE

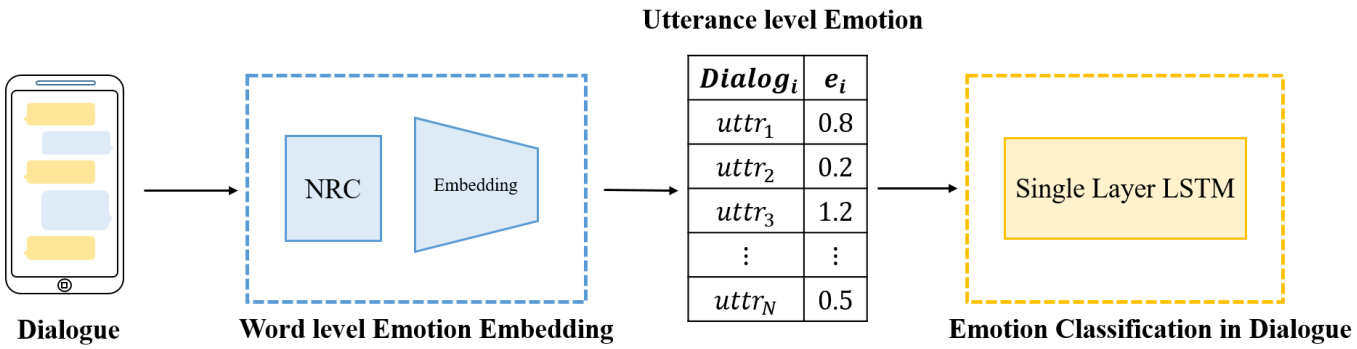

Fig. 2. An overview of proposed model.

A. Word Embedding

The word embedding represents words as continuous vectors based on distributional hypothesis [10]. Among training model, the probabilistic feedforward neural network language model (NNLM) has been proposed in [11]. However, the feedforward NNLM has certain limitations such as the need to specify the context length and the huge complexity per training example. From these reasons, the recurrent neural network based language model has been proposed in [12]. Still it has needed a lot of computation on large amounts of data. Thus, the continuous bag-of-words and continuous skip-gram model have been proposed in [13]. The architectures of these two models are similar to the feedforward NNLM, but have better performance than the earlier work [11], [12]. These models have been a popular method in recent years. In this paper, we modify the continuous skip-gram model to obtain emotion vectors.

B. Text-based Emotion Recognition

The studies on text-based emotion recognition are mainly divided into three categories: keyword-based, learning-based, and hybrid recommendation approaches [14]. Recent researches focus on the learning-based methods. Among the methods, the convolutional neural network (CNN) for sentence classification has a great performance [15]. This study is used for extracting sentence information. As we mentioned, it is hard to recognize the emotion of each utterance in dialogue since it is needed to consider the contextual information. To solve this problem, the contextual LSTM architecture is proposed to measure the inter-dependencies and relations of utterances in dialogue [16]. In this paper, we will conduct experiments the CNN model and the contextual LSTM architecture as baselines.

III. Emotion Classifier

The goal of our model is to classify an emotion of an utterance in dialogue.

A. Overview

An overview of our model is given in Fig. 2. Our model has three mechanisms: **First,** since the word in the same utterance can have similar emotions, we need to embed the emotion as word-level based on the semi-supervised learning. **Second,** we can obtain vectors which represent utterance's emotion through the element-wise summation operator. **Third,** we train a single layer LSTM to classify the emotion of utterance in dialogue.

In the training process, two main parts of our proposed model are trained separately: word-level Emotion Embedding and Emotion Classification in Dialogue. In the inference process, the dialogue is fed into our model to classify the emotion of utterance in dialogue.

B. word-level Emotion Embedding

An utterance is composed of words. To classify an emotion of utterance, it is required to understand emotions of words consists of an utterance. According to the utterance, even the same word can have different emotions. For example, in the following sentences "I love you" and "I hate you", the word "you" which is in "I love you" is more closer to "joy" among the Ekman's six basic emotions. But, the word "you" which is in "I hate you" is more closer to "anger" or "disgust" among the Ekman's six basic emotions. Therefore, we should consider that words in the same utterance have a similar emotions.

The main idea is that co-occurrence words in the same utterance have a similar emotions based on the distributional hypothesis [10]. Thus, we will modify the continuous skip-gram model to represent emotions of words as vectors. Unlike the existing model, our model is trained by the semi-supervised learning. Since it is the semi-supervised learning, we require the labeled data. For labeling emotions for each word, we refer the NRC Emotion Lexicon. The details of the NRC Emotion Lexicon is described in Section IV.

As shown in Fig. 3, a input word w_i is a word in a input utterance $uttr_i$ of length n.

$$uttr_i = \{w_1, w_2, ..., w_n\} \tag{1}$$

The word w_i is encoded using 1-of-V encoding, where V is size of the vocabulary. A weight matrix W has a $V \times D$ dimensions; $W \in \mathbb{R}^{V \times D}$. The input word w_i is projected by the weight matrix W. The encoded vector $enc(w_i)$ with D dimensions represent 1-of-V encoding vector w_i as the

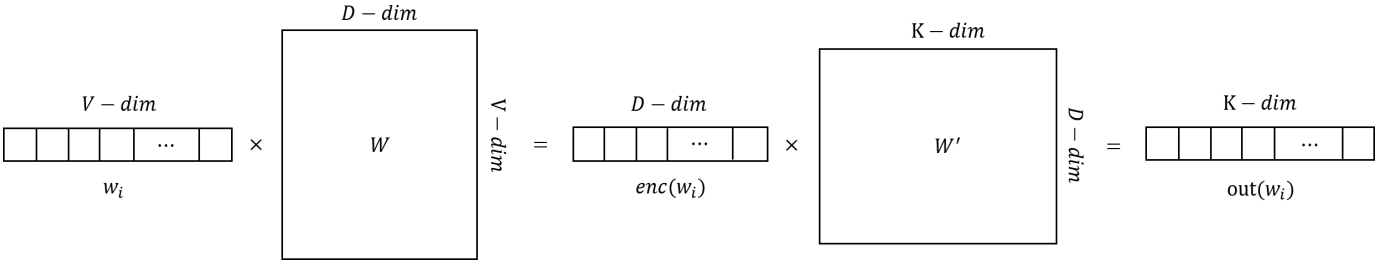

Fig. 3. The word-level emotion embedding architecture.

continuous vector. The result of calculating $enc(w_i)$ with the weight matrix W' is the output vector $out(w_i)$. A weight matrix W' has a $D \times K$ dimensions; $W \in \mathbb{R}^{D \times K}$ where K is the number of emotion label. Then, the predicted output vector $out(w_i)$ can be computed with the expected output vector.

For training this embedding model, we have to make pairs of the input and the expected output. Since this architecture is a slight variant of the skip-gram model, we should choose the maximum distance of the words based on the central word. We should only select the central word which is in the NRC Emotion Lexicon. After selecting the central word, the context words are labeled with the same emotion of the central word as shown in Fig. 1. Through the semi-supervise learning, we can represent the emotion of word as a continuous vector in vector space. For example, if the word "beautiful" in Fig. 1 is not in the NRC Emotion Lexicon, the word "beautiful" will be represented as the emotion "joy" in the continuous vector space.

C. Utterance-level Emotion Representation

From the pre-trained vector, we can get an emotion of an utterance. Let a i-th utterance of length n is represented as (1) where n is not fixed variable. Let $e(w_i)$ is the pre-trained vector which was applied to the word-level emotion embedding. The emotion of the i-th sentence is represented as

$$e(uttr_i) = e(w_1) + e(w_2) + ... + e(w_n) \quad (2)$$

where + is the element-wise summation operator. As we mentioned, all of utterances do not have the same length. For this reason, we use the summation operator not the concatenation operator. Obtained vectors $e(uttr_i)$ will be used to classify emotions in dialogue.

D. Emotion Classification in Dialogue

We will train a single layer LSTM-based classification network on utterance-level emotion vectors obtained from an semi-supervised neural language model. As we described the problem statement, it is important to consider the contextual information in dialogue, such as the emotion flow. In this paper, the emotion flow is regarded as a sequential data. Thus, we will adopt a recurrent neural network (RNN) architecture in the classification model. Let the dialogue consists of several utterances. It is represented as

$$dialogue = \{uttr_1, uttr_2, ..., uttr_C\} \quad (3)$$

where C is not fixed. As shown in Fig. 4, an input $e(uttr_i)$ at time step t is a emotion vectors. At time step t, the predicted output vector and the expected output vector can be computed with a non-linear function such as softmax.

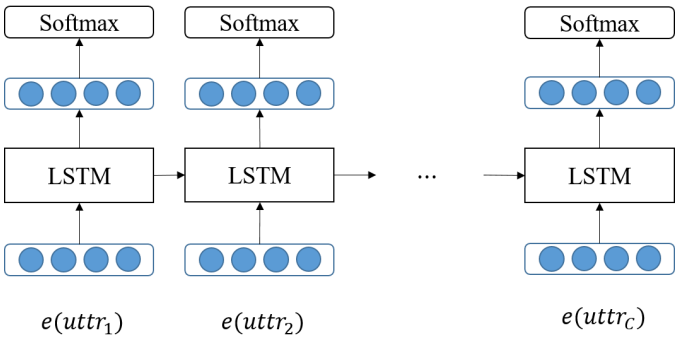

Fig. 4. A single layer LSTM-based classification network.

IV. Datasets

In this studies, it is necessary to require labeled data for training the semi-supervised learning on the word-level emotion vectors. For this, we will use the NRC Emotion Lexicon. Also, we test our model on one benchmark: EmotionLines.

- **EmotionLines:** The first dataset with emotions labeling (six Ekman's basic emotions plus the neutral emotion) on all utterances in each dialogue only based on their textual content. Dialogues are collected from Friends TV scripts and private Facebook messenger dialogues. Each utterance is labeled by 5 Amazon MTurkers [9].

The EmotionLines has annotated utterances with an additional emotion, a non-neutral emotion. Those utterances with more than two different emotions voted were put into the non-neutral emotion. However, since the word-level emotion embedded vectors were labeled with seven emotions, we can not express the non-neutral emotion. Therefore, it is required to regard the non-neutral emotion as others. We think that a current utterance is affected by a previous one. Thus, we consider the non-neutral as the emotion which is labeled on the previous utterance.

- **The NRC Emotion Lexicon:** It is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and

disgust) and two sentiments (negative and positive). The annotations were manually done by crowdsourcing. In this work, we consider eight basic emotions from this lexicon [8].

We will only consider a list of English words with eight basic emotions since our task is about classifying emotions in dialogue. But, there are some problems. In our benchmark, it is labeled with eight emotions (Ekman's six basic emotions plus the neutral and non-neutral emotions). Thus, we will regard anticipation and trust as happiness. Also, we will regard joy as happiness.

V. Experiments Plan

In this section, we will describe our baseline models and evaluation metrics. But, in this paper, we can not detailed descriptions of the training process since we did not experiment.

A. Baselines

As we mentioned, the goal of our model is classifying the emotion in dialogue. Our model combines two tasks. One is sentence classification, another is considering the contextual information in dialogue. On the first one, the CNN model performs remarkably well [15]. On the second one, the contextual LSTM has a good performance [16]. Thus, we will adopt two models as our baselines.

B. Evaluation Metrics

To evaluate our model how well it works, we will adopt emotion accuracy between the expected emotion and the predicted emotion by our model.

VI. Conclusion

In this paper, we proposed the Emotion Classifier in multi-turn conversation. Our model was composed of the word-level emotion embedding, the utterance-level emotion representation, and the emotion classification in dialogue. For the experiments, we plan to use the first dataset with emotions labeling in multi-turn dialogue called EmotionLines. We will evaluate our model with the emotion accuracy.

In our future work, we will process the experiments according to the method described in this paper. After that, we will extend our work to generate emotional response in multi-turn conversation.

Acknowledgment

This research was supported by Korea Electric Power Corporation. (Grant number:R18XA05)

References

[1] T. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE translation journal on magnetics in Japan*, vol. 2, no. 8, pp. 740–741, 1987.

[2] B.-K. Kim, J. Roh, S.-Y. Dong, and S.-Y. Lee, "Hierarchical committee of deep convolutional neural networks for robust facial expression recognition," *Journal on Multimodal User Interfaces*, vol. 10, no. 2, pp. 173–189, 2016.

[3] L. Chao, J. Tao, M. Yang, Y. Li, and Z. Wen, "Audio visual emotion recognition with temporal alignment and perception attention," *arXiv preprint arXiv:1603.08321*, 2016.

[4] H. Lee, Y. S. Choi, S. Lee, and I. Park, "Towards unobtrusive emotion recognition for affective social communication," in *Consumer Communications and Networking Conference (CCNC), 2012 IEEE*. IEEE, 2012, pp. 260–264.

[5] P. Ekman, W. V. Friesen, M. O'sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Heider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti *et al.*, "Universals and cultural differences in the judgments of facial expressions of emotion." *Journal of personality and social psychology*, vol. 53, no. 4, p. 712, 1987.

[6] M. Bouazizi and T. Ohtsuki, "Sentiment analysis: From binary to multi-class classification: A pattern-based approach for multi-class sentiment analysis in twitter," in *Communications (ICC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1–6.

[7] S. Li, L. Huang, R. Wang, and G. Zhou, "Sentence-level emotion classification with label and context dependence," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, vol. 1, 2015, pp. 1045–1053.

[8] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," vol. 29, no. 3, pp. 436–465, 2013.

[9] S.-Y. Chen, C.-C. Hsu, C.-C. Kuo, L.-W. Ku *et al.*, "Emotionlines: An emotion corpus of multi-party conversations," *arXiv preprint arXiv:1802.08379*, 2018.

[10] G. E. Hinton, J. L. McClelland, D. E. Rumelhart *et al.*, *Distributed representations*.

[11] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," *Journal of machine learning research*, vol. 3, no. Feb, pp. 1137–1155, 2003.

[12] T. Mikolov, M. Karafiát, L. Burget, J. Černockỳ, and S. Khudanpur, "Recurrent neural network based language model," in *Eleventh Annual Conference of the International Speech Communication Association*, 2010.

[13] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[14] E. C.-C. Kao, C.-C. Liu, T.-H. Yang, C.-T. Hsieh, and V.-W. Soo, "Towards text-based emotion detection a survey and possible improvements," in *Information Management and Engineering, 2009. ICIME'09. International Conference on*. IEEE, 2009, pp. 70–74.

[15] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[16] S. Ghosh, O. Vinyals, B. Strope, S. Roy, T. Dean, and L. Heck, "Contextual lstm (clstm) models for large scale nlp tasks," *arXiv preprint arXiv:1602.06291*, 2016.