Beyond Single Emotion: Multi-label Approach to Conversational Emotion Recognition

Yujin Kang, Yoon-Sik Cho

Department of Artificial Intelligence, Chung-Ang University, Republic of Korea
{zinzin32, yoonsik}@cau.ac.kr

Abstract

Emotion recognition in conversation (ERC) has been promoted with diverse approaches in the recent years. However, many studies have pointed out that *emotion shift* and *confusing labels* make it difficult for models to distinguish between different emotions. Existing ERC models suffer from these problems when the emotions are forced to be mapped into single label. In this paper, we utilize our strategies for extending single label to multi-labels. We then propose a multi-label classification framework for emotion recognition in conversation (ML-ERC). Specifically, we introduce weighted supervised contrastive learning tailored for multi-label, which can easily applied to previous ERC models. The empirical results on existing task with single label support the efficacy of our approach, which is more effective in the most challenging settings: emotion shift or confusing labels. We also evaluate ML-ERC with the multi-labels we produced to support our contrastive learning scheme.

Introduction

Motivated by the introduction of AI conversational systems, such as chatbots in healthcare, recommendations, and customer services, emotion recognition in conversation (ERC) has attracted increasing attention in recent years. The ERC task aims to identify the emotion at each utterance in a conversation. While consistent improvement is being shown, there still remain challenges for improvement. Previous studies have pointed out that *emotion shift* (Hazarika et al. 2018; Song et al. 2022b; Tu et al. 2023) and *confusing labels* (Ghosal et al. 2019; Ishiwatari et al. 2020; Lee 2022) are the causes that make the ERC task difficult (Majumder et al. 2019; Li et al. 2021; Shen et al. 2021; Yang et al. 2022; Qin et al. 2023). An emotion shift in ERC occurs when the emotions of the same speaker change during the speaker's consecutive utterances. Confusing emotion, where two similar emotions cannot be distinguished within an utterance, is another challenge in ERC. Recent studies have proposed models to address these issues by introducing emotion shift detection module (Gao et al. 2022), using curriculum learning (Yang et al. 2022) to better handle emotion shift, and constructing grayscale labels (Lee 2022).

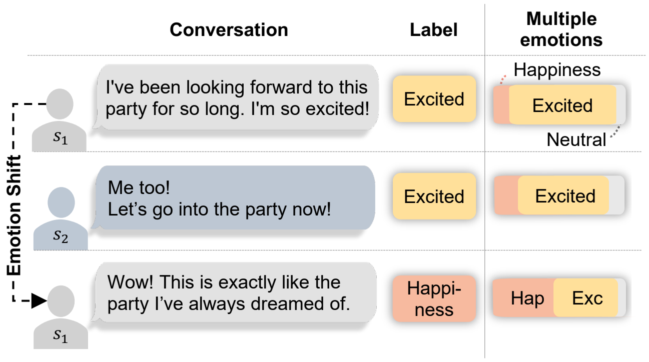
Conversation
Label
Multiple emotions
s_1
I've been looking forward to this party for so long. I'm so excited!
Excited
Happiness
Excited
Neutral
s_2
Me too! Let's go into the party now!
Excited
Excited
Emotion Shift
s_1
Wow! This is exactly like the party I've always dreamed of.
Happiness
Hap
Exc

Figure 1: Example of multiple emotions in each utterance within a conversation. Specifically, emotion shift frequently triggers multiple emotions, making it difficult to understand with only a single emotion label.

These two problems, emotion shift and confusing labels, arise from the practice of annotating a single emotion label to an utterance overlooking that an utterance can encompass multiple emotions. As illustrated in Figure 1, each utterance within a conversation is assigned to a single label. Although an utterance can encompass multiple emotions (right), only the most intense emotion is retained and other emotions are discarded. In fact, organizing emotions in 2D or circular arrangements that originated from a study in the 1950s (Schlosberg 1952) has become more common. The studies from psychology (Russell 1980; Mikels et al. 2005) use valence-arousal 2D emotion space to describe emotions. From these perspectives, restricting each utterance to one emotion label is oversimplified. Mikels et al. (2005) pointed out that emotions are often blended and expressed through behaviors or utterances, which has motivated a new approach to capturing mixed-emotions. Regardless of these studies, however, the emotion recognition task in ERC still remains in predicting the single label.

To address the aforementioned challenges, we propose a model called Multi-Label classification for Emotion Recognition in Conversation (**ML-ERC**). Although all previous ERC models attempted to predict a *single*-label for each utterance, we switch this same ERC task to *multi*-label (emotion) prediction. Since multi-label annotation requires significant time and efforts, we thus propose a pseudo multi-label assignment strategy *without additional cost* for multi-emotion labels. The self-annotation scheme is devised based on the studies in human emotions (Kuppens, Allen, and Sheeber 2010; Koval et al. 2015) and inductive reasoning. Specifically, when assigning the additional emotion labels, we conversely make use of the *emotion shift*, which previously had a negative impact.

We employ the supervised contrastive learning (SupCon) (Khosla et al. 2020) scheme for our multi-label ERC task. However, given that SupCon is originally designed for *single* supervisory signals, it cannot be directly extended to multi-label settings. To bridge this gap, we introduce a novel multi-label weighted supervised contrastive loss, *MulWCL*, designed to better account for multi-label tasks. This objective makes multi-label classification more effective, and surprisingly it also contributes to performance improvements in single-label classification by mitigating the challenges from emotion shift and confusing labels prevalent in ERC field. Our contribution is three-fold.

- We approach ERC by utilizing a multi-label to address the two problems: emotion shift and confusing labels. For this new approach, pseudo multi-labeling scheme for multi-label is introduced.
- Our ML-ERC incorporates weighted supervised contrastive loss to consider the characteristics of multi-label classification, and employs a soft multi-labeling method within the module to facilitate the training process.
- We conduct extensive experiments to verify the effectiveness of our proposed model. We integrate our multi-label scheme into existing single label ERC models, and show how our objective improves all of the existing baseline models.

Related Work

Emotion recognition in conversation Previous works on emotion recognition in *textual* conversation can be summarized into three methods: Recurrence-based, Graph-based, and Knowledge-enhanced methods. Recurrence-based works (Hazarika et al. 2018; Majumder et al. 2019; Hu et al. 2023) consider utterances as sequential data. Graph-based models (Ghosal et al. 2019; Ishiwatari et al. 2020; Shen et al. 2021) represent the relationship of an utterance using a graph. The knowledge-enhanced models improve the performance of ERC by associating external knowledge (Ghosal et al. 2020; Zhu et al. 2021; Lee and Lee 2022). There are also methods other than these three. Yang et al. (2022) improve performance by applying hybrid curriculum learning. Gao et al. (2022) propose a multi-task learning framework that employs emotion shift detection as an auxiliary task and ERC as the main task. Lee (2022) attempts to understand emotion using the grayscale label. Zhang et al. (2023) mimic human thinking through the use of prompts and paraphrasing. Recently, several works (Song et al. 2022a; Yang et al. 2023; Hu et al. 2023; Kang and Cho 2024) utilize contrastive learning to effectively learn emotion.

Multi-label classification Multi-label classification has gained continuous attention in the field of NLP due to its practical applications (Nam et al. 2017; Zhang et al. 2021). Particularly, multi-label emotion classification (ER) has seen active research progress (Alhuzali and Ananiadou 2021; Lin et al. 2023) with the release of datasets, which provide multi-label emotion annotations for tweets (Mohammad et al. 2018) and multimodal language (Zadeh et al. 2018). However, they are quite different from ERC.

Model	original F1	w/o ES	only ES
DialogueRNN	62.75	69.2(+6.45)	47.5(-15.25)
GloVe bcLSTM	61.90	-	52.37(-9.53)
TODKAT	62.60	64.62(+2.02)	56.24(-6.36)
TODKAT+HCL	63.03	67.01(+3.98)	56.91(-6.12)

Table 1: Impact of emotion shift on the performance of ERC. IEMOCAP dataset was used for the evaluation. The 'only ES', 'w/o ES' represent utterances with emotion-shift and without emotion-shift, respectively.

Since dialogue involves many complex factors, multi-label annotation on each utterance in ERC is challenging and requires human-intensive resources. MEISD dataset (Firdaus et al. 2020) marks a pioneering effort to introduce multi-label annotations in ERC. Zhao et al. (2022) propose multi-label emotional dialogue in chinese. Despite their efforts, studies in ERC still remain in single classifications with single-labeled datasets. To the best of our knowledge, yet there are no models in the literature for multi-label classification in ERC. In this study, we introduce a pseudo-label strategy that can build multi-labels from existing datasets. We further introduce a multi-label classification model which well reflects the nature of emotions in ERC.

Multi-label contrastive learning Supervised contrastive learning (SupCon) (Khosla et al. 2020) has contributed to performance improvements in NLP (Gunel et al. 2020; Lin et al. 2022). However, it is not readily applicable to multi-label instances since this learning approach assumes that the sample has single label. Consequently, there are few investigations for contrastive learning for multi-label in the NLP field. Su, Wang, and Dai (2022) introduce multi-label contrastive learning based on the label similarity of multi-label instances. Lin et al. (2023) propose five contrastive losses designed for multi-label text classification. These works primarily focus on using the label correlations across instances. In this work, we consider a multi-label contrastive learning approach from two perspectives: the class-level and the instance-level.

Proposed Approach

Motivation for Multi-label ERC

Many ERC studies have discussed *emotion shift* and *confusing emotion*. Here we investigate how these phenomena affect classification performance in ERC. The first is the issue of emotion shift, which refers to the transition of emotions. Table 1 summarizes the experimental results from several studies (Majumder et al. 2019; Ghosal et al. 2021; Yang et al. 2022), revealing the influence of emotion shift on performance. According to Table 1, F1 score always increases when emotion shifts are taken out from the test data. When

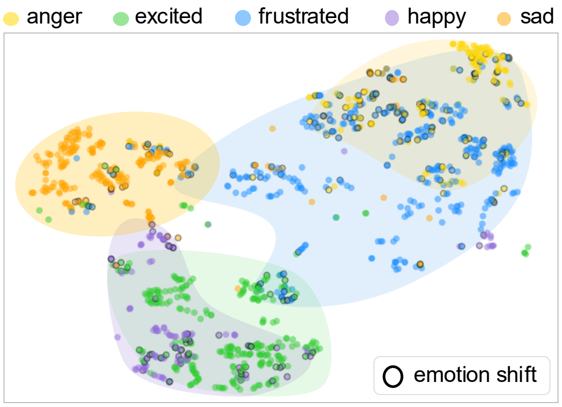

Figure 2: The t-SNE visualization on test set of IEMOCAP dataset. Data points marked with a black border indicates samples where emotion shift occured. The marking denotes that each label dominates this space.

performing evaluations on emotion shifts only, we observe considerable performance degradation, which is up to 15%. The second key challenge is the confusing labels. Previous methods (Majumder et al. 2019; Xie et al. 2021; Shen et al. 2021; Li et al. 2021) often failed to distinguish subtle differences between certain emotions, such as excited-happy and anger-frustrated. Due to the vague boundaries of confusing emotions, many models struggle to classify these emotions.

Generating pseudo multi-emotion labels In this study, we tackle the challenges in ERC with multi-label classification. In multi-label classification, each instance can be associated with multiple labels, which means, in ERC, each utterance can be annotated with multiple emotion labels. However, the current benchmark datasets are labeled with single emotions. To fill this gap, we propose a self annotation scheme for generating pseudo multi-emotion labels without incurring additional costs.

For the automatic annotation strategy, we specifically employ *emotion inertia* (Kuppens, Allen, and Sheeber 2010; Koval et al. 2015), a concept from psychological theory describing the resistance to changes in emotions. When an emotional change occurs, the preceding emotion's influence results in the persistence of that emotional state. Based on emotion inertia, the pseudo labels are aligned to *emotion shift* allowing two emotions (old and new) to coexist at every emotional shift. To support our approach, we visualize the embeddings of utterances from IEMOCAP, a benchmark dataset for ERC, using RoBERTa-large as the embedding module. As illustrated in Figure 2, data points with emotion shifts are primarily located within or near the overlapping regions of each emotion. These observations suggest that emotion-shift data may simultaneously encompass multiple emotions.

Our pseudo multi-label scheme can generate multiple labels for each utterance from the existing dataset with single labels without any human effort for labeling. If the same speaker's present utterance and the previous utterance have different labels in conversation, and none of which are *neutral*, we target the current utterance for multi-label annotation. When the specified conditions are satisfied, we generate pseudo multi-labels by aggregating two emotions from the previous and current utterances of the same speaker; otherwise, we keep the annotation as single. We also provides additional examples applying our scheme to the existing dataset in Appendix B (see Figure S1 and Table S1).

Multi-label ERC Model

Problem definition Each ERC dataset consists of multiple independent conversations, where each conversation is a sequence of utterances attached to speaker and emotion: $C = \{(u_i, s_i, y_i)\}_{i=1}^{N}$. Here s_i, y_i represent the speaker and label of u_i and N denotes the number of utterances in a conversation. When target utterance (u_t, s_t) and its context $\{(u_i, s_i)\}_{i=1}^{t-1}$ are given, the goal of single-label classification of ERC is to predict the emotion label (y_t) of target u_t in the predefined label set $K = \{k_1, k_2, \dots, k_{|K|}\}$.

Motivated by our findings, we approach the problem from the point of view of multi-label classification task by associating an utterance to multiple labels. We restructure the dataset by adding multi-hot label $\mathbf{y}^{\text{pseudo}}$ to the existing data and expand C to $C = \{(u_i, s_i, y_i, \mathbf{y}_i^{\text{pseudo}})\}_{i=1}^{N}$. We define $\mathbf{y}_i^{\text{pseudo}} = \{k_i^1, k_i^2, \dots, k_i^{|K|}\} \in \mathbb{Z}_2^{|K|}$, where each emotion $k^j \in \{0, 1\}$. The value $k_i^j = 1$ indicates that the i-th utterance has emotion k_j. Throughout this paper, we rename the current emotion label as the **main emotion** and define the additional emotion labels from the multi-label settings as the **sub emotions**.

Embedding module We bring RoBERTa-Large (Liu et al. 2019), a pre-trained language model (PLM), as an embedding module. For each utterance, we prepend its respective speaker and concatenate it with the context of the current utterance. A special token [CLS], which reflects context information, is placed at the beginning of this sequence. In embedding stage, the input and output of u_i are as follows:

$$\text{RoBERTa}([\text{CLS}], s_1, u_1, \dots, s_i, u_i) = \mathbf{h}_i \tag{1}$$

,where $\mathbf{h}_i \in \mathbb{R}^{1 \times d_h}$ is the embedding of [CLS] token of u_i in the last hidden layer.

Multi-label prediction module In multi-label setting, more than one labels can coexist in one sample. Therefore, given the $\mathbf{h}_i$, embedding of u_i, our model predicts the multi-label of u_i following Equation 2 – 4.

$$\mathbf{z}_i = \text{Linear}(\mathbf{h}_i), \tag{2}$$

$$\tilde{\mathbf{z}}_i = \tanh(\mathbf{z}_i), \tag{3}$$

$$\mathbf{o}_{ij} = \begin{cases} 1 & \text{if } \tilde{z}_{ij} > \text{mean}(\tilde{\mathbf{z}}_i) \\ 0 & \text{otherwise} \end{cases}, j \in \{1, ..., |K|\} \tag{4}$$

, where $\mathbf{z}_i \in \mathbb{R}^{1 \times |K|}$ and $\mathbf{o}_{ij}$ represent the prediction for label k_j of data u_i. Here we use calibrated threshold (Hou et al. 2021) by taking the mean of each component in $\tilde{\mathbf{z}}$.

ML-ERC learning objectives Supervised contrastive learning (SupCon) (Khosla et al. 2020) pulls data points with same label closer to the anchor while repelling negative samples from the anchor. However, SupCon cannot be directly

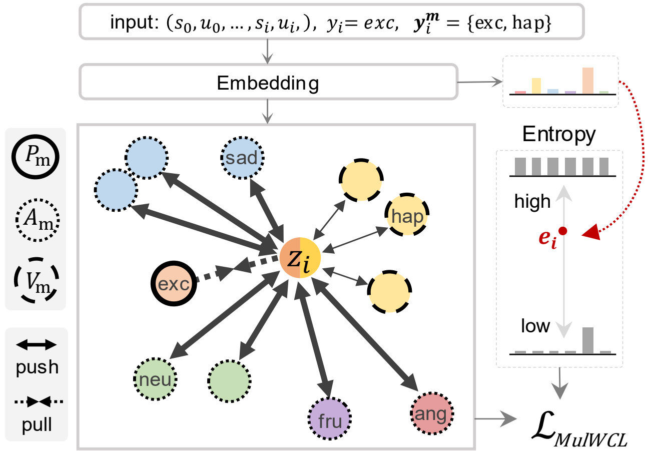
input: $(s_0, u_0, \ldots, s_i, u_i,)$, y_i= exc, $\mathbf{y}_i^{\boldsymbol{m}}$ = {exc, hap}
Embedding
Entropy
high
e_i
low
P_m
A_m
V_m
push
pull
sad
hap
z_i
exc
neu
fru
ang
$\mathcal{L}_{MulWCL}$

Figure 3: Framework of multi-label weighted supervised contrastive loss (*MulWCL*). The thickness of the lines indicating the intensity of the repelling force and different colors of the sample represent different labels.

used in the multi-label setting for two reasons. First, finding the positive and negatives pairs with multi-label vector is too complex as the size of label set increases. Second, as the multi-label can coexist within a sample, it causes blurring effect (Lin et al. 2023), where the multiple emotions cause overlapping positive and negative pairs. To address this, we introduce a multi-label weighted supervised contrastive loss, *MulWCL*, designed specifically for a task with multi-labels.

For the first challenge, we redefine the positive and negative sets in our settings. Given the extensive range of possible label combinations in multi-label settings, finding a positive pair with fully matching labels is challenging. Therefore, we define the positive set $P(i)$ as instances sharing a main emotion with the anchor. For negative set of sample i, we introduce two distinct sets: $A_m(i)$ and $V_m(i)$.

$$P_m(i) = \{p | y_p = y_i, p \neq i\}. \tag{5}$$

$$\mathbf{m}_i = \{k_j | k_i^j \neq 0, k_i^j \in \mathbf{y}_i^{\text{pseudo}}\},$$

$$A_m(i) = \{a | y_a \notin \mathbf{m}_i, a \neq i\}, \tag{6}$$

$$V_m(i) = \{v | y_v \in (\mathbf{m}_i - y_i), v \neq i\}. \tag{7}$$

$A_m(i)$ is the true negative set comprised of samples whose labels do not belong to the multi-label of sample i. $V_m(i)$ is a set comprising samples not matching the main label of sample i but aligning with one of sub emotions of sample i. While the vanilla SupCon pushes V_m away with the same force as it does samples in true negative set A_m, we reduce the weight of repulsion in V_m. The intuition behind this is that any pair of samples from multi-label set V_m can have looser repulsion than the samples drawn from the true negative set A_m in the latent space. When a sample takes multiple emotions, an emotion can be both positive and negative to a given sample, which causes the second challenge, *blurring effect*. Since we categorize the samples into a positive set and two negative sets, our MulWCL ensures that there is no overlap between the three sets.

We design two weights from different perspectives: class-level and instance-level. The samples exhibiting multiple emotions pose challenges in representing each emotion clearly, compared to samples characterized by a single label. Thus, we apply weights tailored to the multi-label context to achieve better distinctions between emotions. We design the class-level weighted score for the multi-label set V_m by reflecting the similarity between the anchor's main emotion and sub emotion in V_m. We adjust the repelling force, reducing it as the similarity between labels increases. The calculations for both positive and negative scores are presented in Equations 8 and 9.

$$\mathcal{P}_{\text{multi}}(i) = \sum_{p \in P_m(i)} \exp(\text{sim}(\mathbf{h}_i, \mathbf{h}_p)/\tau), \tag{8}$$

$$\mathcal{N}_{\text{multi}}(i) = \sum_{a \in A_m(i)} \exp(\text{sim}(\mathbf{h}_i, \mathbf{h}_a)/\tau) \tag{9}$$

$$+ \sum_{v \in V_m(i)} \underbrace{(1 - \frac{(\text{sim}(r_i, r_v) + 1)}{2})}_{\text{class weight}} \cdot \exp(\text{sim}(\mathbf{h}_i, \mathbf{h}_v)/\tau),$$

where the $\text{sim}(\cdot)$ calculates similarity between two samples and we use a cosine similarity. τ is a temperature parameter. r indicates label representation in Valence-Arousal dimension obtained from previous studies (Yang et al. 2022, 2023). Further details regarding the relation between emotions can be found in the Appendix B.

Furthermore, we calculate the instance-level weighted score, leveraging entropy measure. We try to capture patterns more from the samples with distinct emotions while paying less attention to the samples with mixed emotions. Thus, we apply small weights to the samples with high entropy.

$$e_i = \frac{-\sum_{j=1}^{|K|} \sigma(\mathbf{z}_i)_j \log \sigma(\mathbf{z}_i)_j}{\log |K|}. \tag{10}$$

$$\mathcal{L}_{\text{MulWCL}}(i) = \underbrace{(1 - e_i)}_{\text{instance weight}} \left(\frac{-1}{|P_m(i)|} \log \frac{\mathcal{P}_{\text{multi}}(i)}{\mathcal{N}_{\text{multi}}(i)} \right). \tag{11}$$

Entropy is calculated in Equation 10 and σ is softmax function. The entropy inverted to serve as weights. These instance-wise weights adjust the attraction within positive pairs and the repulsion within negative pairs. We highlight that our loss function, $\mathcal{L}_{\text{MulWCL}}$, applies both class-wise and instance-wise weights through label similarity in Equation 9 and entropy in Equation 11. Finally, the overall loss we optimize is presented below.

$$\mathcal{L}_{\text{ML-ERC}} = \mathcal{L}_{bce} + \alpha \mathcal{L}_{\text{MulWCL}}, \tag{12}$$

where $\mathcal{L}_{bce}$ is the binary cross-entropy loss and α is hyperparameter that controls the effect of our weighted multi-label loss.

Soft multi-labeling As multi-label annotation is time-consuming and expensive, we assign pseudo multi-label on the data where emotion shifts occur. However, we speculate that there still could be more utterances with multiple emotions that our pseudo labeling scheme has missed. We additionally introduce *soft multi-labels* annotated to the potential

Algorithm 1: Learning procedure of ML-ERC for each batch $\mathcal{B}$ at each epoch. Once the model runs several iterations, we conduct soft-labeling.

```
Input: $\mathcal{B} = \{(u_i, s_i, y_i, \mathbf{y}_i^{\text{pseudo}}, \mathbf{p}_i^{\text{soft}} = \mathbf{y}_i^{\text{pseudo}})\}_{i=1}^{N_b}$;
       $K$ = single label set;
Output: $\mathcal{L}_{\text{ML-ERC}}$;
        $\mathcal{B}_{\text{new}}$ = batch updated with soft-labeling

$\mathcal{B}_{\text{new}} = []$
for $i = 1, \dots, N_b$ do
    $\mathbf{o}_i = \{o_{ij} = 0\}_{j=1}^{|K|}$
    $\mathbf{h}_i$ = RoBERTa($u_i, s_i$, context)
    $\tilde{\mathbf{z}}_i$ = Normalize($\mathbf{h}_i$) ;                      // Eq 2, 3
    for $j = 1, \dots, |K|$ do
        if $mean(\tilde{\mathbf{z}}_i) < \tilde{\mathbf{z}}_{ij}$ then
            $o_{ij} = 1$
        end
    end
    calculate entropy $e_i$ ;                                   // Eq 10
    $\mathbf{p}_{\text{new}} = \mathbf{y}_i^{\text{pseudo}}$
    if $count(\{x \neq 0 \mid x \in \mathbf{y}_i^{pseudo}\})$ == 1 then
        if $\gamma < e_i$ then
            $\mathbf{p}_{\text{new}} = \mathbf{o}_i$
        end
    end
    $\mathcal{B}_{\text{new}}$.append(($u_i, s_i, y_i, \mathbf{y}_i^{\text{pseudo}}, \mathbf{p}_i^{\text{soft}} = \mathbf{p}_{\text{new}}$))
end
$\mathcal{L}_{\text{ML-ERC}} = \mathcal{L}_{\text{BCE}}(\mathbf{p}^{\text{soft}}, \mathbf{o}) + \alpha\mathcal{L}_{\text{MulWCL}}(\mathbf{h}, \mathbf{y}, \mathbf{p}^{\text{soft}})$
$\mathcal{B} = \mathcal{B}_{\text{new}}$ ;    // update batch $\mathcal{B}$ to $\mathcal{B}_{\text{new}}$ for next
 iteration
```

utterances using the embeddings learned from the model. To enhance the efficacy of soft-labeling, we employ the entropy calculated in Equation 10 as criterion for soft-label generation. This approach aims to avoid incorrectly applying multi-labeling to data that is unlikely to exhibit multiple emotions.

$$\mathbf{p}_i^{\text{soft}} = \begin{cases} \mathbf{o}_i & \text{if } \left(\sum_{j=1}^{|K|} \mathbf{1}(y_j^{\text{pseudo}} \neq 0) = 1\right) \text{ and } (\gamma < e_i) \\ \mathbf{y}_i^{\text{pseudo}} & \text{otherwise} \end{cases} \tag{13}$$

, where $\mathbf{p}_i^{\text{soft}}$ and $\mathbf{o}_i$ represent the soft multi-label and the multi-label prediction of sample i, respectively. The $\mathbf{1}(\cdot)$ is an indicator function, and γ is a hyperparameter that controls the threshold for soft-label assignment.

If the entropy of the data exceeds a pre-defined value (γ), we additionally assign the multi-labels $\mathbf{o}_i$ predicted for u_i in Equation 4 as its soft-labels. These soft-labels are then used as pseudo multi-labels for the corresponding data in the next training epoch. It is worth noting that soft-labeling is applied exclusively during the training phase and is not used during the inference stage. To ensure the quality of the soft-labels, we initially train our model for a sufficient number of iterations until the model becomes reliable, and integrate the soft-label strategy into the training process. This kind of strategy can be seen in other works (Wang et al. 2020). Our ML-ERC is outlined in the Algorithm 1.

Experimental Settings

Data We conduct experiments on three benchmark ERC datasets annotated with single labels. The statistics for each dataset are provided in Table S4 in Appendix F. For the ERC task, only text scripts are being used. *EmoryNLP* (Zahiri and Choi 2018) is labeled with joyful, mad, neutral, peaceful, powerful, scared, and sad from the Feeling wheel (Willcox 1982). *MELD* (Poria et al. 2019) is a multi-modal dataset with a label set that includes anger, disgust, fear, joy, neutral, surprise, and sadness. *IEMOCAP* (Busso et al. 2008) is a dyadic multimodal dataset with labels including excited, neutral, frustrated, sadness, happiness, and anger.

Baselines We bring representative methods in ERC : recurrence-based `DialogueRNN` (Majumder et al. 2019), graph-based `DAG-ERC` (Shen et al. 2021), knowledge-based `CoMPM` (Lee and Lee 2022), and PLM based `MPLP` (Zhang et al. 2023). Furthermore, we implement the `RoBERTa-large` model (Liu et al. 2019) by adding a classification layer to the top of the embedding.

Training setup To train our ML-ERC method, we set the learning rate, the number of batch sizes and epochs are 1e-6, 16 and 30, respectively. We fix τ in Eq 8, 9 to 0.05. For α in Eq 12, we search the parameter using the validation set. We set α to 0.7 for Emorynlp, 0.1 for MELD, and 0.4 for IEMOCAP. All experiments are performed on an Nvidia RTX A6000 GPU. Further details on parameters are provided in Appendix E.

Experimental Results and Analysis

Our approach using multi-label to ERC benchmark dataset is new in the ERC literature. We therefore evaluate our model in two ways to verify the effectiveness of approach: *single-label* classification and *multi-label* classification.

Results of Single-label Classification

Previous ERC methods have suffer from the problems of *emotion shift* and *confusing labels*, which are the motivation of our work. Here, we verify how our multi-label scheme mitigates the aforementioned challenges.

Implementation and evaluation details We capture the final representation of the data and the loss values directly from baseline ERC models. The embeddings generated by the ERC models are then leveraged to compute the multi-label loss in Equation 12. We integrate the multi-label loss values ($\mathcal{L}_{\text{ML-ERC}}$) with the original loss from ERC model ($\mathcal{L}_{\text{ERC}}$) through the hyperparameter β, which is set to 0.5, and optimize a given model. Thus, the final loss for single-label classification is expressed as $\mathcal{L} = \beta\mathcal{L}_{\text{ERC}} + (1 - \beta)\mathcal{L}_{\text{ML-ERC}}$. We maintain the hyperparameter settings as defined by the original models without conducting any additional tuning when integrating our objective for fair comparison. Following previous ERC studies, we use the weighted F1-score as our evaluation metric.

Effect of multi-label objectives Table 2 shows the efficacy of multi-label objective. All of the results are reproduced using the original code. The ERC models attain a performance boost through our loss. We consistently achieve

Base Model	Loss	Dataset EMORY	 MELD	 IEMOCAP
RoBERTa	$\mathcal{L}_{\text{ERC}}$	33.08	64.27	63.58
	$+\mathcal{L}_{\text{ML-ERC}}$	**35.74**	**65.15**	**63.96**
DialogueRNN	$\mathcal{L}_{\text{ERC}}$	37.44	57.03	62.75
	$+\ \mathcal{L}_{\text{ML-ERC}}$	**38.18**	**57.51**	**63.56**
DAG-ERC	$\mathcal{L}_{\text{ERC}}$	38.85	63.38	67.07
	$+\ \mathcal{L}_{\text{ML-ERC}}$	**39.02**	**63.58**	**68.12**
CoMPM	$\mathcal{L}_{\text{ERC}}$	36.20	64.87	66.47
	$+\ \mathcal{L}_{\text{ML-ERC}}$	**37.35**	**65.90**	**68.00**
MPLP	$\mathcal{L}_{\text{ERC}}$	-	65.45	65.03
	$+\ \mathcal{L}_{\text{ML-ERC}}$	-	**65.93**	**66.09**

Table 2: Experiment results in single-label classification.

Base Model	Loss	Dataset EMORY	 MELD	 IEMOCAP
RoBERTa	$\mathcal{L}_{\text{ERC}}$	29.17	51.37	45.01
	$+\ \mathcal{L}_{\text{ML-ERC}}$	**31.37**	**55.15**	**48.20**
DialogueRNN	$\mathcal{L}_{\text{ERC}}$	36.81	**41.87**	46.86
	$+\ \mathcal{L}_{\text{ML-ERC}}$	**37.97**	41.82	**53.41**
DAG-ERC	$\mathcal{L}_{\text{ERC}}$	**35.81**	50.88	63.25
	$+\ \mathcal{L}_{\text{ML-ERC}}$	35.01	**50.95**	**63.78**
CoMPM	$\mathcal{L}_{\text{ERC}}$	22.50	50.03	**46.14**
	$+\ \mathcal{L}_{\text{ML-ERC}}$	**29.79**	**53.45**	45.60
MPLP	$\mathcal{L}_{\text{ERC}}$	-	52.45	49.34
	$+\ \mathcal{L}_{\text{ML-ERC}}$	-	**55.28**	**53.78**

Table 3: Experiment results in emotion shift data.

Base Model	Objective	Confusing labels (↓) Peaceful - Happy	 Powerful - Happy	 Sad – Fear	 Angry - Frustrated	 Excited - Happy
RoBERTa	$\mathcal{L}_{\text{ERC}}$	34.14	38.96	28.67	26.39	40.77
	$+\mathcal{L}_{\text{ML-ERC}}$	**33.60 (-0.54)**	**32.37 (-6.59)**	**24.81 (-3.86)**	**23.24 (-3.15)**	**24.44 (-16.33)**
CoMPM	$\mathcal{L}_{\text{ERC}}$	33.25	34.32	30.89	28.20	29.10
	$+\mathcal{L}_{\text{ML-ERC}}$	**26.94 (-6.31)**	**33.91 (-0.41)**	**24.97 (-5.92)**	**22.16 (-6.04)**	**25.66 (-3.44)**
DAG-ERC	$\mathcal{L}_{\text{ERC}}$	**24.07**	35.94	25.52	19.16	33.78
	$+\mathcal{L}_{\text{ML-ERC}}$	24.76 (+0.69)	**32.15 (-3.79)**	**25.42 (-0.10)**	**18.95 (-0.21)**	**33.01 (-0.77)**

Table 4: The misclassified rate (**lower the better**) as confusing labels on ERC datasets. We select confusing labels, which have a similarity of 0.6 or higher between two emotions in Figure S2 in Appendix.

performance improvements on all the baselines across three datasets. It is worth noting that we inject a multi-label perspective to ERC models without additional training data. We speculate that our method takes advantage of the rich information overlooked by previous single-label approaches.

Performance on emotion shift We perform single-label classification on selected test data which involves emotion shift, which is presented in Table 3. Compared to the performance of the full dataset in Table 2, Table 3 shows a significant performance drop ranging from 3% to 20% on the emotion shift data. These results show that traditional ERC methods are vulnerable to classify samples with emotion shift. However, incorporating our multi-label objective into the training of ERC models enhances performance across most evaluations, while some results slightly underperform the baselines.

Performance on confusing labels Table 4 shows the proportion which incorrectly classified as confusing emotions compared with the true label. Confusing emotions are closely positioned in embedding space, highly co-occurring. Training with single labels forces multiple emotions into one emotion, bringing similar emotions closer in the embedding space. Our MulWCL effectively reduces the overlapping areas between these confusing emotions. Thus, models trained with our method classify emotions in similar spaces better (less confused) than original model.

Results of Multi-label Classification

In previous section, we demonstrated that our multi-label approach allviates the issues in ERC and achieves performance improvements. In this section, we evaluate the multi-label classification performance of our proposed framework.

Implementation and evaluation details In ERC, research on multi-label classification is rudimentary. Thus, we apply our multi-label prediction module to single-label baseline models. We calculate the embedding values for each utterance through the ERC models, and then predict the multi-labels for each embedding using a calibrated threshold in equation 4. [1] As evaluation metrics, we choose weighted F1 scores and macro F1 score. We additionally use metrics used in other multi-label tasks, such as AUC (Area Under the Curve) and hamming loss.

Multi-label classification performance The results are reported in Table 5, where we compare ML-ERC against other multi-label performance extended from baselines. Our ML-ERC outperforms the multi-label performances of baseline models. The ERC models trained on single-label datasets tend to overfocus on tracking the strongest signal while disregarding the probabilities of other emotional signals. Thus, extending the single-label ERC to multi-label by using a calibrated threshold still exhibits limited performance on multi-label classification.

Ablation study We first strip independently two components, MulWCL and soft multi-labeling in Table 6. The results demonstrate that omitting MulWCL consistently leads to performance degradation, highlighting its vital importance in ML-ERC. The absence of soft multi-labeling leads

[1] Note that we maintain the training process and hyperparameter setting of each ERC model without modification.

Model	EmoryNLP				MELD				IEMOCAP			
	M-F1(↑)	W-F1(↑)	AUC(↑)	HL(↓)	M-F1(↑)	W-F1(↑)	AUC(↑)	HL(↓)	M-F1(↑)	W-F1(↑)	AUC(↑)	HL(↓)
DialogueRNN	32.80	40.62	0.583	**0.2388**	34.37	56.61	0.612	0.1622	61.77	62.95	0.758	0.1440
DAG-ERC	36.47	41.33	0.586	0.3647	39.13	55.91	0.649	0.3038	58.34	57.97	0.772	0.2814
CoMPM	37.76	39.74	0.617	0.4018	39.80	55.12	0.714	0.2720	57.06	58.19	0.808	0.2803
MPLP	-	-	-	-	41.71	55.49	0.691	0.2466	59.39	59.99	0.811	0.2421
ML-ERC	**38.69**	**41.56**	**0.630**	0.2535	**50.03**	**63.01**	**0.718**	**0.1250**	**68.58**	**69.17**	**0.815**	**0.1312**

Table 5: Experiment results in multi-label classification. M-F1, W-F1, and HL represent macro-F1 (%), weighted-F1 (%) and hamming loss, respectively. Higher macro-F1, weighted-F1, and AUC, along with lower hamming loss, indicate better performance. We highlight the best performance among the main results in bold. All results are reproduced using their respective official codebase.

Model	EmoryNLP					MELD					IEMOCAP				
	S-F1(↑)	M-F1(↑)	W-F1(↑)	AUC(↑)	HL(↓)	S-F1(↑)	M-F1(↑)	W-F1(↑)	AUC(↑)	HL(↓)	S-F1(↑)	M-F1(↑)	W-F1(↑)	AUC(↑)	HL(↓)
ML-ERC	35.74	38.69	41.56	0.630	0.253	65.15	50.03	63.01	0.718	0.125	63.96	68.58	69.17	0.815	0.131
w/o MulWCL	34.85	37.72	40.70	0.625	0.253	64.21	48.64	62.87	0.715	0.130	62.14	66.81	67.97	0.807	0.131
w/o Soft-label	34.75	38.05	40.88	0.624	0.256	64.16	50.16	63.28	0.721	0.126	63.29	68.42	68.96	0.817	0.128
w/o Both	33.12	36.63	39.38	0.620	0.267	63.65	48.64	62.40	0.717	0.125	63.13	65.35	67.23	0.796	0.137

Table 6: Ablation study. S-F1 is weighted F1-score. S-F1 is the result of single-label classification, whereas M-F1, W-F1, AUC, and HL are the outcomes of multi-label classification.

Loss	EmoryNLP				MELD				IEMOCAP			
	M-F1(↑)	W-F1(↑)	AUC(↑)	HL(↓)	M-F1(↑)	W-F1(↑)	AUC(↑)	HL(↓)	M-F1(↑)	W-F1(↑)	AUC(↑)	HL(↓)
BCE	34.57	37.12	0.597	0.2888	48.60	62.52	0.716	0.1256	65.41	66.28	0.816	0.1339
+ SupCon	36.63	39.38	0.620	0.2675	48.64	62.40	0.717	0.1258	65.35	67.23	0.796	0.1373
+ SCL	37.03	39.95	0.617	0.2632	48.97	62.76	0.712	0.1251	67.24	68.52	0.807	0.1325
+ JSCL	37.86	40.50	0.619	0.2674	49.88	62.28	0.714	**0.1248**	67.35	68.16	0.820	0.1339
+ JSPCL	33.46	35.77	0.589	0.2949	49.07	62.56	0.716	0.1261	66.18	66.98	**0.821**	0.1478
+ SLCL	17.27	23.23	0.503	0.3134	39.19	58.63	0.668	0.1496	61.12	62.11	0.792	0.1869
+ ICL	34.24	36.56	0.592	0.3058	49.02	63.10	0.720	0.1313	62.34	63.34	0.794	0.1698
+ MulWCL	**38.05**	**40.88**	**0.624**	**0.2566**	**50.16**	**63.28**	**0.721**	0.1260	**68.42**	**68.96**	0.817	**0.1288**

Table 7: Comparisons against multi-label contrastive losses. Bold score indicates the best performance, and underlined score indicates the second-best performance in each evaluation setting. All of the results in baselines are implemented with method outlined in the original paper.

to reduced performance on both the EmoryNLP and IEMOCAP datasets. These results reveal that the two modules designed for multi-label classification effectively enhance performance in both multi-label and single-label classification. For extended experiments on MulWCL and soft multi-labeling, see Appendix C.

Comparisons against multi-label contrastive losses To better demonstrate the effectiveness of our multi-label contrastive loss, we replace it with other learning objectives and compare the performances. Lin et al. (2023) has introduced five contrastive losses, `SCL`, `JSCL`, `JSPCL`, `SLCL`, and `ICL` tailored for multi-label contexts (details provided in Appendix D). We exclude pseudo labeling from our model to strictly verify the effect of multi-label contrastive loss of ours. In Table 7, we find that MulWCL outperforms other multi-label contrastive losses in most metrics.

Conclusion

In this paper, we propose a novel ERC model for Multi-Label classification for Emotion Recognition in Conversation (*ML-ERC*) to tackle problems when the emotions are constrained to a single label. ML-ERC employs a novel weighted supervised contrastive learning (*MulWCL*) to obtain better representative embedding and a soft-labeling method to facilitate multi-label classification. The experimental results show that ML-ERC not only exhibits superior performance in multi-label classification but also achieves a performance boost for all ERC baselines by effectively mitigating the ERC challenges.

Acknowledgements

This work was partly supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00419201) and Institute of

Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. 2021-0-01341, Artificial Intelligence Graduate School Program of Chung-Ang Univ.).

References

Alhuzali, H.; and Ananiadou, S. 2021. SpanEmo: Casting Multi-label Emotion Classification as Span-prediction. In Merlo, P.; Tiedemann, J.; and Tsarfaty, R., eds., *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 1573–1584. Online: Association for Computational Linguistics.

Busso, C.; Bulut, M.; Lee, C.-C.; Kazemzadeh, A.; Mower Provost, E.; Kim, S.; Chang, J.; Lee, S.; and Narayanan, S. 2008. IEMOCAP: Interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 42: 335–359.

Firdaus, M.; Chauhan, H.; Ekbal, A.; and Bhattacharyya, P. 2020. 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*, 4441–4453.

Gao, Q.; Cao, B.; Guan, X.; Gu, T.; Bao, X.; Wu, J.; Liu, B.; and Cao, J. 2022. Emotion recognition in conversations with emotion shift detection based on multi-task learning. *Knowledge-Based Systems*, 248: 108861.

Ghosal, D.; Majumder, N.; Gelbukh, A.; Mihalcea, R.; and Poria, S. 2020. COSMIC: COmmonSense knowledge for eMotion Identification in Conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2470–2481.

Ghosal, D.; Majumder, N.; Mihalcea, R.; and Poria, S. 2021. Exploring the role of context in utterance-level emotion, act and intent classification in conversations: An empirical study. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 1435–1449.

Ghosal, D.; Majumder, N.; Poria, S.; Chhaya, N.; and Gelbukh, A. 2019. DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 154–164.

Gunel, B.; Du, J.; Conneau, A.; and Stoyanov, V. 2020. Supervised Contrastive Learning for Pre-trained Language Model Fine-tuning. In *International Conference on Learning Representations*.

Hazarika, D.; Poria, S.; Zadeh, A.; Cambria, E.; Morency, L.-P.; and Zimmermann, R. 2018. Conversational memory network for emotion recognition in dyadic dialogue videos. In *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting*, volume 2018, 2122. NIH Public Access.

Hou, Y.; Lai, Y.; Wu, Y.; Che, W.; and Liu, T. 2021. Few-shot learning for multi-label intent detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 13036–13044.

Hu, D.; Bao, Y.; Wei, L.; Zhou, W.; and Hu, S. 2023. Supervised Adversarial Contrastive Learning for Emotion Recognition in Conversations. *arXiv preprint arXiv:2306.01505*.

Ishiwatari, T.; Yasuda, Y.; Miyazaki, T.; and Goto, J. 2020. Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 7360–7370.

Kang, Y.; and Cho, Y.-S. 2024. Improving Contrastive Learning in Emotion Recognition in Conversation via Data Augmentation and Decoupled Neutral Emotion. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2194–2208.

Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised contrastive learning. *Advances in Neural Information Processing Systems*, 33: 18661–18673.

Koval, P.; Brose, A.; Pe, M. L.; Houben, M.; Erbas, Y.; Champagne, D.; and Kuppens, P. 2015. Emotional inertia and external events: The roles of exposure, reactivity, and recovery. *Emotion*, 15(5): 625.

Kuppens, P.; Allen, N. B.; and Sheeber, L. B. 2010. Emotional inertia and psychological maladjustment. *Psychological science*, 21(7): 984–991.

Lee, J. 2022. The Emotion is Not One-hot Encoding: Learning with Grayscale Label for Emotion Recognition in Conversation. In *Proc. Interspeech 2022*, 141–145.

Lee, J.; and Lee, W. 2022. CoMPM: Context Modeling with Speaker's Pre-trained Memory Tracking for Emotion Recognition in Conversation. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 5669–5679.

Li, J.; Lin, Z.; Fu, P.; and Wang, W. 2021. Past, present, and future: Conversational emotion recognition through structural modeling of psychological knowledge. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, 1204–1214.

Lin, H.; Ma, J.; Chen, L.; Yang, Z.; Cheng, M.; and Guang, C. 2022. Detect Rumors in Microblog Posts for Low-Resource Domains via Adversarial Contrastive Learning. In *Findings of the Association for Computational Linguistics: NAACL 2022*, 2543–2556.

Lin, N.; Qin, G.; Wang, G.; Zhou, D.; and Yang, A. 2023. An Effective Deployment of Contrastive Learning in Multi-label Text Classification. In *Findings of the Association for Computational Linguistics: ACL 2023*, 8730–8744.

Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; and Stoyanov, V. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Majumder, N.; Poria, S.; Hazarika, D.; Mihalcea, R.; Gelbukh, A.; and Cambria, E. 2019. Dialoguernn: An attentive rnn for emotion detection in conversations. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 6818–6825.

Mikels, J. A.; Fredrickson, B. L.; Larkin, G. R.; Lindberg, C. M.; Maglio, S. J.; and Reuter-Lorenz, P. A. 2005. Emotional category data on images from the International Affective Picture System. *Behavior research methods*, 37(4): 626–630.

Mohammad, S.; Bravo-Marquez, F.; Salameh, M.; and Kiritchenko, S. 2018. SemEval-2018 Task 1: Affect in Tweets. In Apidianaki, M.; Mohammad, S. M.; May, J.; Shutova, E.; Bethard, S.; and Carpuat, M., eds., *Proceedings of the 12th International Workshop on Semantic Evaluation*, 1–17. New Orleans, Louisiana: Association for Computational Linguistics.

Nam, J.; Loza Mencía, E.; Kim, H. J.; and Fürnkranz, J. 2017. Maximizing subset accuracy with recurrent neural networks in multi-label classification. *Advances in neural information processing systems*, 30.

Poria, S.; Hazarika, D.; Majumder, N.; Naik, G.; Cambria, E.; and Mihalcea, R. 2019. MELD: A Multimodal Multi-Party Dataset for Emotion Recognition in Conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics.

Qin, X.; Wu, Z.; Zhang, T.; Li, Y.; Luan, J.; Wang, B.; Wang, L.; and Cui, J. 2023. BERT-ERC: Fine-tuning BERT is enough for emotion recognition in conversation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 13492–13500.

Russell, J. A. 1980. A circumplex model of affect. *Journal of personality and social psychology*, 39(6): 1161.

Schlosberg, H. 1952. The description of facial expressions in terms of two dimensions. *Journal of experimental psychology*, 44(4): 229.

Shen, W.; Wu, S.; Yang, Y.; and Quan, X. 2021. Directed Acyclic Graph Network for Conversational Emotion Recognition. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 1551–1560. Online: Association for Computational Linguistics.

Song, X.; Huang, L.; Xue, H.; and Hu, S. 2022a. Supervised Prototypical Contrastive Learning for Emotion Recognition in Conversation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, 5197–5206.

Song, X.; Zang, L.; Zhang, R.; Hu, S.; and Huang, L. 2022b. Emotionflow: Capture the Dialogue Level Emotion Transitions. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 8542–8546.

Su, X.; Wang, R.; and Dai, X. 2022. Contrastive Learning-Enhanced Nearest Neighbor Mechanism for Multi-Label Text Classification. In Muresan, S.; Nakov, P.; and Villavicencio, A., eds., *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 672–679. Dublin, Ireland: Association for Computational Linguistics.

Tu, G.; Jing, R.; Liang, B.; Yang, M.; Wong, K.-F.; and Xu, R. 2023. A Training-Free Debiasing Framework with Counterfactual Reasoning for Conversational Emotion Detection. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 15639–15650.

Wang, L.; Liu, Y.; Qin, C.; Sun, G.; and Fu, Y. 2020. Dual relation semi-supervised multi-label learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 6227–6234.

Willcox, G. 1982. The feeling wheel: A tool for expanding awareness of emotions and increasing spontaneity and intimacy. *Transactional Analysis Journal*, 12(4): 274–276.

Xie, Y.; Yang, K.; Sun, C.-J.; Liu, B.; and Ji, Z. 2021. Knowledge-Interactive Network with Sentiment Polarity Intensity-Aware Multi-Task Learning for Emotion Recognition in Conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, 2879–2889.

Yang, K.; Zhang, T.; Alhuzali, H.; and Ananiadou, S. 2023. Cluster-level contrastive learning for emotion recognition in conversations. *IEEE Transactions on Affective Computing*.

Yang, L.; Shen, Y.; Mao, Y.; and Cai, L. 2022. Hybrid curriculum learning for emotion recognition in conversation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 11595–11603.

Zadeh, A. B.; Liang, P. P.; Poria, S.; Cambria, E.; and Morency, L.-P. 2018. Multimodal language analysis in the wild: Cmu-mosei dataset and interpretable dynamic fusion graph. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2236–2246.

Zahiri, S. M.; and Choi, J. D. 2018. Emotion detection on tv show transcripts with sequence-based convolutional neural networks. In *Workshops at the thirty-second aaai conference on artificial intelligence*.

Zhang, T.; Chen, Z.; Zhong, M.; and Qian, T. 2023. Mimicking the thinking process for emotion recognition in conversation with prompts and paraphrasing. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 6299–6307.

Zhang, X.; Zhang, Q.-W.; Yan, Z.; Liu, R.; and Cao, Y. 2021. Enhancing Label Correlation Feedback in Multi-Label Text Classification via Multi-Task Learning. In Zong, C.; Xia, F.; Li, W.; and Navigli, R., eds., *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, 1190–1200. Online: Association for Computational Linguistics.

Zhao, J.; Zhang, T.; Hu, J.; Liu, Y.; Jin, Q.; Wang, X.; and Li, H. 2022. M3ED: Multi-modal Multi-scene Multi-label Emotional Dialogue Database. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 5699–5710.

Zhu, L.; Pergola, G.; Gui, L.; Zhou, D.; and He, Y. 2021. Topic-driven and knowledge-aware transformer for dialogue emotion detection. *arXiv preprint arXiv:2106.01071*.