# LREC-COLING 2024

#### **CEPT: a Contrast-Enhanced Prompt-Tuning Framework for Emotion Recognition in Conversation**

Qingqing Gao<sup>1</sup>, Jiuxin Cao<sup>1</sup>, Biwei Cao<sup>1</sup>, Xin Guan<sup>1</sup>, Bo Liu<sup>2</sup>

<sup>1</sup>School of Cyber Science and Engineering, Southeast University, Nanjing, China <sup>2</sup>School of Computer Science and Engineering, Southeast University, Nanjing, China



## Background

So, C press	asey, how are things going with the photos for the kit? [Neutral]
7	Yeah, I've been meaning to talk to you about that. I m
	need to ask for an extension on that deadline. [Neutra
You' thing	ve had over a month to get this finalized! Why are s delayed? [Anger]
	Well, the thing is, we ran into a lot of problems [Sad

A example from DailyDialog dataset

#### Motivation



Emotion distributions in three benchmark datasets: MELD, DailyDialog and IEMOCAP

## Methodology: Overview



The architecture of the Contrast-Enhanced Prompt-Tuning (CEPT) Framework

### Methodology: Contributions

- Bridging the gap between Pre-trained Language Model's Masked Language Modeling task and the Emotion Recognition in Conversation task via prompt-tuning, mitigating the issue of insufficient information on minority emotions;
- Mining more information from emotion labels via Supervised Contrastive Learning, reducing biased predictions for the common emotions.

### Methodology: Prompt construction



- Hard words ("Context: " and "Target: ") indicate the boundaries of context and target utterance.
  - Soft words ("The emotion is: ") guide the Pre-trained
    Language Model to generate the emotion of the target utterance.

## Methodology: Label mapping

Category	Label words
Anger / Angry	anger, frustration, irritation, hostility
Disgust	disgust, revulsion, nausea, aversion
Fear	fear, anxiety, apprehension, nervousness
Happiness / Joy / Happy	joy, happiness, contentment, satisfaction
Sadness / Sad	sadness, loss, disappointment, grief
Surprise	surprise, astonishment, amazement
Excited	excitement, thrill, exhilaration
Frustrated	frustration, disappointment, dissatisfaction, annoyance
Neutral / no_ emotion	neutral / no_ emotion

Mapping between original emotion category and label words

#### Methodology: Label mapping

The probability that the emotion category of utterance  $u_i$  is  $e_j$ , is calculated based on all  $e_j$ 's label words  $EW_j = \{ew_1, ew_2, \dots, ew_{k_i}\}$ :

$$p(e_j|u_i) = \sum_{ew_j \in EW_j} p([MASK] = ew_j|I_i)$$

For a batch with *D* utterances, the loss of MLM generation is calculated using cross-entropy loss as follows:

$$L_{Gen} = -\frac{1}{D} \sum_{i=1}^{D} [\mathbf{y}_i] log(\mathbf{P}_i)$$

## Methodology: Supervised contrastive learning

We stack the predicted emotion probability distribution vectors of all utterances within a batch as a matrix  $H_b$  and make a copy of it as  $H'_b$ , whose gradient is detached to ensure the parameter optimization is stable.

The vectors used for computing the Supervised Contrastive Learning loss are denoted as  $H^m = [H_b, H'_b] = \{h_1^m, h_2^m, \dots, h_D^m, h_{D+1}^m, \dots, h_{2D}^m\}$ 

## Methodology: Supervised contrastive learning

The calculation formula for Supervised Contrastive Learning:

$$SIM(h_i^m, h_p^m) = \log \frac{\exp(h_i^m \cdot h_p^m / \tau)}{\sum_{a \in A(i)} \exp(h_i^m \cdot h_a^m / \tau)}$$

$$L_{SCL} = \sum_{i=1}^{2D} \frac{-1}{|P(i)|} \sum_{p \in P(i)} SIM(h_i^m, h_p^m)$$

## Methodology: Training

The loss that guides the training process of CEPT :

 $L = (1 - \alpha)L_{Gen} + \alpha L_{SCL}$ 

## Experiments: Dataset

Dataset	# of <b>C</b>	Conversa	tions	# of	Utteran	ces
Dataset	train	valid	test	train	valid	test
MELD	1038	114	280	9989	1109	2610
DailyDialog	11118	1000	1000	87170	8069	7740
IEMOCAP	108	12	31	5163	647	1623

The statistics of the datasets

#### **Experiments: Evaluation Metrics**

- MELD : Weighted average F1 (Weighted-F1);
- DailyDialog: Micro average F1 (Micro-F1) and exclude the "Neutral" labels;
- ➤ IEMOCAP: Weighted average F1 (Weighted-F1).

#### Experiments: Overall performance

Dataset	MELD	DD	IEMOCAP
DialogueRNN	57.10	50.27	62.75
ESD-ERC	62.15	57.44	-
DialogXL	62.67	54.93	65.94
DAG-ERC	63.42	59.33*	68.03
CoG-BART	64.90	56.29	66.18
CISPER	66.08	-	-
SPCL-CL-ERC	66.96*	-	69.74*
RoBERTa-ERC	64.61	52.87	51.65
CEPT	67.51	61.52	70.53

Performance Comparison with the baseline and state-of-the-art methods

#### Experiments: Performance on each emotion category

Emotion	Anger	Disgust	Fear	Joy	Neutral	Sadness	Surprise
(Number)	(345)	(68)	(50)	(402)	(1256)	(208)	(281)
DialogueRNN	42.26	00.00	00.00	52.79	76.11	21.59	46.78
ESD-ERC	48.40	00.00	00.00	59.49	79.21	27.33	58.41
DialogXL	49.93	00.00	00.00	61.25	78.55	33.16	57.56
DAG-ERC	49.17	30.09*	26.98	60.25	77.22	36.57	58.22
CoG-BART	47.34	19.35	30.00*	62.15	79.47	43.40	58.41
CISPER	56.80	23.53	28.89	61.37	80.53	38.83	56.69
SPCL-CL-ERC	56.91*	27.66	25.88	63.34*	80.57*	42.01	58.98*
RoBERTa-ERC	50.74	24.00	9.84	61.89	79.53	39.13	57.28
CEPT	57.06	32.32	31.58	64.53	80.73	42.39*	59.02

Performance Comparison in each emotion category of different methods on MELD

## Experiments: Ablation study

	Dromot	ampt Label mapping		MELD	DailyDialog	IEMOCAP
5	Frompt		SUL	(Weighted-F1)	(micro-F1)	(Weighted-F1)
_				67.51	61.52	70.53
	×	$\overline{\mathbf{v}}$		65.75 (↓1.76)	55.18 (↓6.34)	51.85 (↓18.68)
		×		66.29 (↓1.22 )	57.66 (↓3.86)	68.31 (↓2.22)
		$\checkmark$	×	66.02 (↓1.49)	60.96 (↓0.56)	67.88 (↓2.65)

The ablation results of CEPT on three datasets

## Experiments: Prompt analysis

	Template
(1)	"Context: " $\oplus C_i \oplus$ "Target: " $\oplus s(u_i) \oplus$ ": " $\oplus u_i \oplus$ "The emotion is: " $\oplus \{"mask"\}$
(2)	$\{"soft"\} \oplus C_i \oplus \{"soft"\} \oplus s(u_i) \oplus ": " \oplus u_i \oplus \{"soft"\} \{"soft"\} \{"soft"\} \oplus \{"mask"\}$
(3)	$\{"soft": "Context:"\} \oplus C_i \oplus \{"soft": "Target:"\} \oplus s(u_i) \oplus ":" \oplus u_i$
(0)	$\oplus \{"soft": "The emotion is:"\} \oplus \{"mask"\}$
(4)	"Context: " $\oplus C_i \oplus$ "Target: " $\oplus s(u_i) \oplus$ ": " $\oplus u_i \oplus \{$ "soft"} $\{$ "soft"} $\{$ "soft"} \oplus \{"mask"}
(5)	"Context: " $\oplus C_i \oplus$ "Target: " $\oplus s(u_i) \oplus$ ": " $\oplus u_i \oplus \{$ "soft": "The emotion is:" $\} \oplus \{$ "mask" $\}$

	MELD	DailyDialog	IEMOCAP
(1)	64.97	60.60	65.34
(2)	65.54	60.06	66.91
(3)	66.08	59.48	64.73
(4)	64.27	60.72	65.32
(5)	67.51	61.52	70.53

The five prompt templates using five different strategies and the performance of CEPT with different prompt templates

#### Experiments: Parameters analysis

P	3	4	5	6	7	8	P	3	4	5	6	7	8	P	3	4	5	6	7	8
10	67.04	66.64	67.59	66.81	66.71	66.87	10	58.61	58.64	61.00	58.91	60.64	59.99	 10	68.50	69.31	68.30	67.59	68.19	67.97
9	66.93	66.89	66.81	66.71	67.21	66.52	9	59.16	60.44	58.76	59.14	60.28	59.54	9	68.21	67.70	70.53	67.82	68.27	68.09
8	66.74	67.81	67.32	67.05	66.68	67.14	8	60.13	60.36	59.63	59.80	61.52	60.38	8	69.03	69.04	69.96	69.21	69.24	69.69
7	66.67	67.53	67.51	67.11	66.65	66.14	7	60.87	60.66	59.30	61.10	59.80	59.45	7	67.50	66.68	68.74	70.35	67.81	67.32
6	66.37	67.33	66.96	66.81	66.71	66.81	6	59.59	60.89	59.06	60.00	59.68	60.20	6	66.74	68.78	67.77	67.03	67.88	67.50
5	66.18	67.03	67.48	67.16	66.26	65.98	5	60.44	59.30	59.85	58.47	57.86	59.14	5	66.51	67.83	66.88	67.42	67.22	67.79

The performance of CEPT with different context window sizes on MELD, DailyDialog and IEMOCAP

#### Experiments: Parameters analysis

$\alpha$	0.1	0.15	0.2	0.25	0.3	0.35	0.4	
MELD	67.51	66.95	67.32	67.25	67.09	67.05	67.28	
DailyDialog	59.68	61.34	59.65	60.36	59.18	59.96	61.37	
IEMOCAP	68.94	68.86	68.33	68.96	69.45	68.63	70.53	
$\alpha$	0.45	0.5	0.55	0.6	0.65	0.7	0.75	
MELD	67.29	66.88	67.34	66.96	67.23	66.86	67.09	
DailyDialog	59.25	61.49	59.95	60.89	61.52	60.25	59.94	
IEMOCAP	69.54	69.21	69.10	69.93	69.18	68.63	67.56	







The performance of CEPT with different Supervised Contrastive Learning loss weights on MELD, DailyDialog and IEMOCAP

## Experiments: Case Study

Speaker	Utterance	Ground truth	CEPT	CISPER
Ross	Hi.	Neutral	Neutral	Neutral
Rachel	Rachel.	Neutral	Neutral	Neutral
Ross	Rachel! Well, you-you're not at home, you're-you're- you're right here.	Surprise	Surprise	Anger
Rachel	Yeah I know, and I bet you thought it would be weird. But it's not!	Joy	Joy	Neutral
Ross	Okay. So well I'll umm, I'll have her home by mid- night.	Neutral	Neutral	Neutral

Case 1 from MELD with the ground-truth emotion labels and the predicted labels from CEPT and CISPER.

## Experiments: Case Study

Speaker	Utterance	Ground truth	CEPT	SPCL-CL-ERC
Joey	Here. I need to borrow some moisturizer.	Neutral	Neutral	Neutral
Monica	For what?	Neutral	Neutral	Neutral
Joey	Whaddya think? Today's the big day!	Joy	Joy	Joy
Monica	Oh my God. Okay, go into the bathroom, use	Disgust	Disgust	Surprise
	whatever you want, just don't ever tell me what			
	you did in there.			
Joey	Thank you!	Joy	Joy	Joy

Case 2 from MELD with the ground-truth emotion labels and the predicted labels from CEPT and SPCL-CL-ERC.

#### Conclusions

- A context-aware mixed prompt template and a label mapping strategy for prompt-tuning of the Pre-trained Language Model;
- Supervised Contrastive Learning for extracting more information from the labels;
- Excellent overall performance and outstanding performance in recognizing both minority emotions and common emotions.

## LREC-COLING 2024

## Thank you!

Qingqing Gao, Jiuxin Cao, Biwei Cao, Xin Guan, Bo Liu

#### Email: qingqing\_gao@seu.edu.cn

