Supervised Prototypical Contrastive Learning for Emotion Recognition in Conversation

Xiaohui Song[⋆†‡], **Longtao Huang**[⋆], **Hui Xue**[⋆] and **Songlin Hu**[†‡]
[†]Institute of Information Engineering, Chinese Academy of Sciences
[‡]School of Cyber Security, University of Chinese Academy of Sciences
[⋆]Alibaba Group
{songxiaohui,husonglin}@iie.ac.cn, {kaiyang.hlt, hui.xueh}@alibaba-inc.com

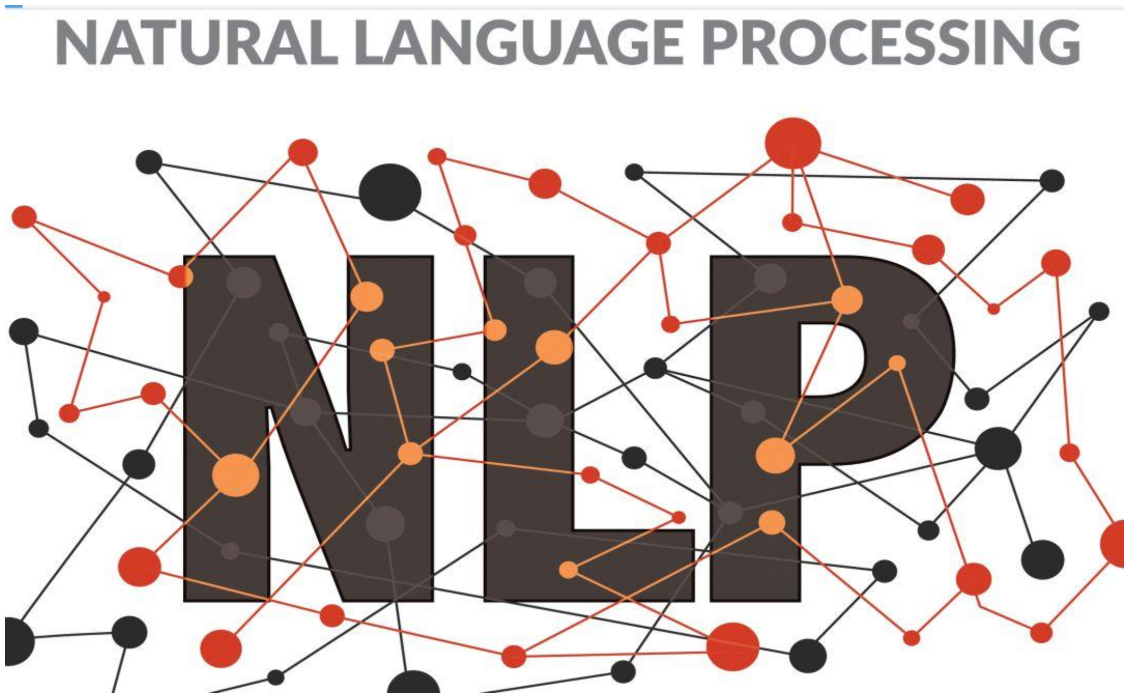
2022.11.17 • ChongQing

2022_EMNLP

Leibniz-Institut für Sozialwissenschaften

Reported by Yidan Liu

https://github.com/caskcsg/SPCL



1.Introduction

2.Method

3.Experiments

Introduction

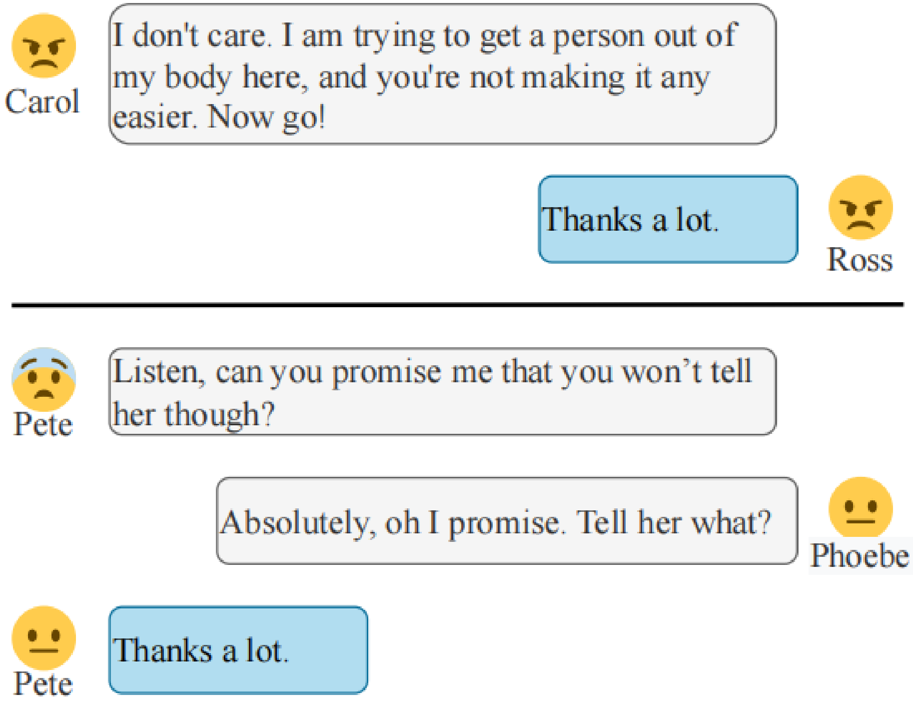


Figure 1: Examples of emotion recognition in conversation. The same utterance "Thanks a lot" can convey different emotions in different contexts.

Even semantically similar utterances, the emotion may vary drastically depending on contexts or speakers.

Introduction

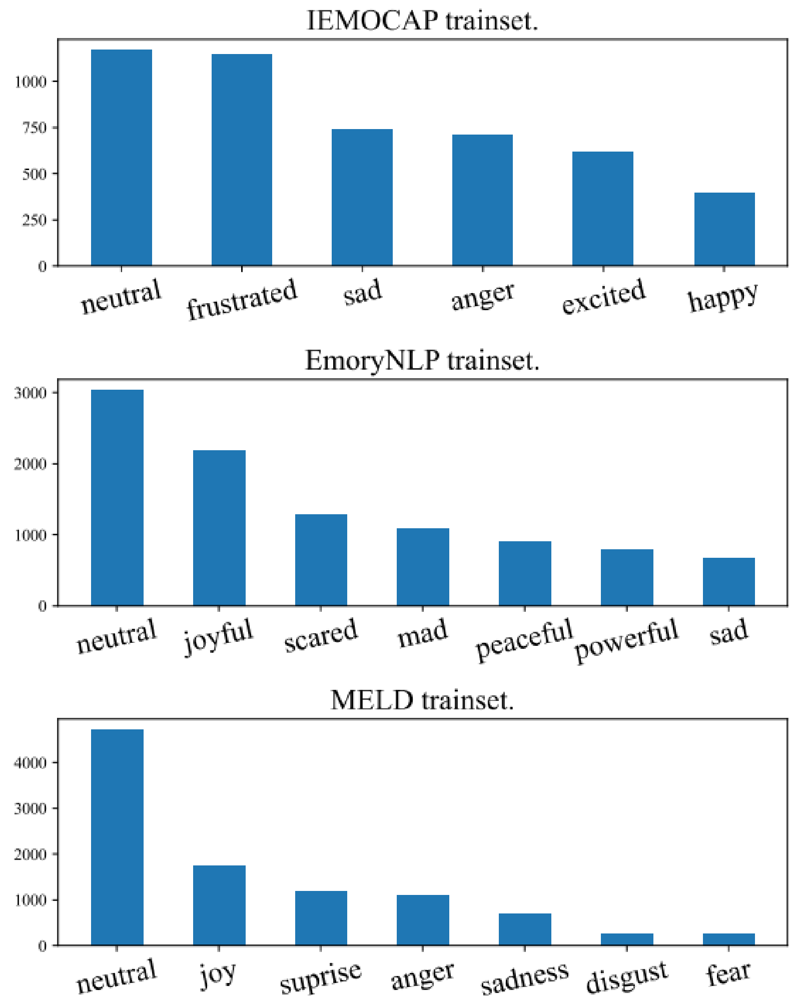


Figure 2: Emotion distributions of the three datasets.

Existing ERC datasets are often class-imbalanced and samples may not be able to meet appropriate positive/negative samples within a mini-batch.

There are some conversations whose textual information is insufficient to distinguish emotions.

Method

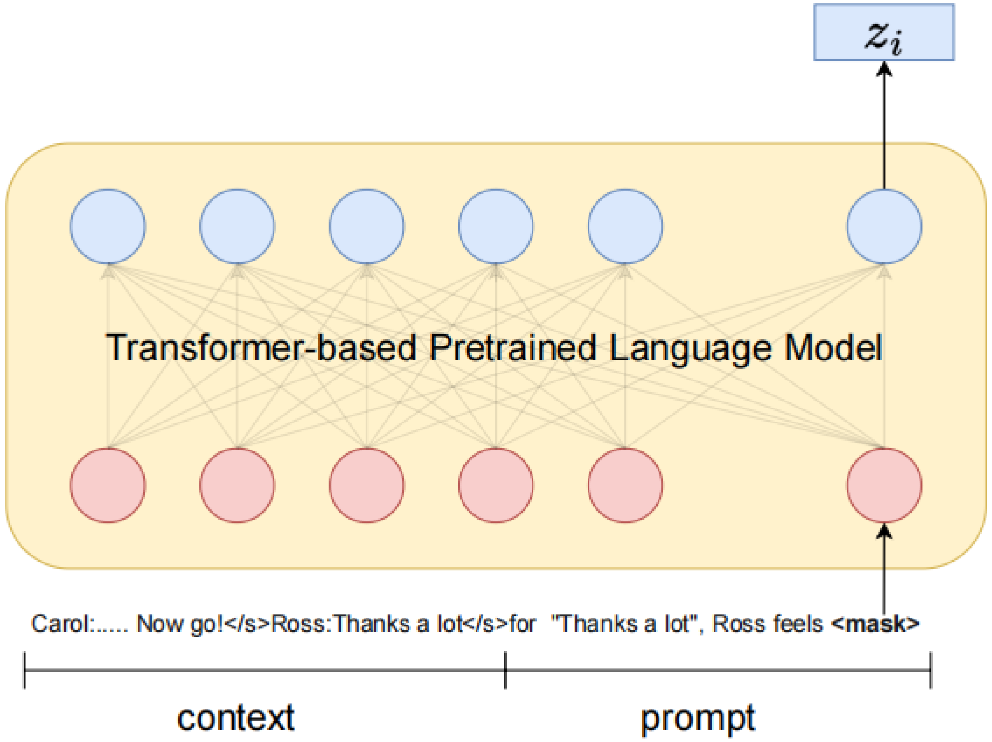


Figure 3: The architecture of our prompt-based context encoder.

all speakers $\mathcal{S}$ label set $\mathcal{E}$

$$[(s_1, u_1), (s_2, u_2), \cdots, (s_N, u_N)],$$

$$C_t = [s_{t-k}, u_{t-k}, s_{t-k+1}, ..., s_t, u_t] \qquad (1)$$

$$P_t = \texttt{for } u_t, s_t \texttt{ fells <mask>} \qquad (2)$$

$$H_t^k = \mathrm{SimCSE}(C_t \oplus P_k) \qquad (3)$$

the embeddings of the special token <mask> from H_t^k as a representation of y_k-th emotion.

Method

Supervised Contrastive Learning

$$I = [z_1, z_2, \cdots, z_N]$$

$$\mathcal{F}(z_i, z_j) = \exp(\mathcal{G}(z_i, z_j)/\tau) \tag{4}$$

$$\mathcal{N}_{sup}(i) = \sum_{z_j \in A(i)} \mathcal{F}(z_i, z_j) \tag{5}$$

$$A(i) \equiv I \backslash \{z_i\}$$

$$\mathcal{P}_{sup}(i) = \sum_{z_p \in P(i)} \mathcal{F}(z_i, z_p) \tag{6}$$

$$\mathcal{L}_i^{sup} = -\log \frac{1}{|P(i)|} \frac{\mathcal{P}_{sup}(i)}{\mathcal{N}_{sup}(i)} \tag{7}$$

Prototypical Contrastive Learning

$$Q_i = [z_1^i, z_2^i, \cdots, z_M^i]$$

$$S_K = \textsc{RandomSelect}(Q_i, K) \tag{8}$$

$$\mathbf{T}_i = \frac{1}{K} \sum_{z_j^i \in S_K, j \in [1 \ldots K]} z_j^i \tag{9}$$

$$\mathcal{N}_{spcl}(i) = \mathcal{N}_{sup}(i) + \sum_{k \in \mathcal{E} \backslash y_i} \mathcal{F}(z_i, \mathbf{T}_k) \tag{10}$$

$$\mathcal{P}_{spcl}(i) = \mathcal{P}_{sup}(i) + \mathcal{F}(z_i, \mathbf{T}_{y_i}) \tag{11}$$

$$\mathcal{L}_i^{spcl} = -\log \left(\frac{1}{|P(i)| + 1} \cdot \frac{\mathcal{P}_{spcl}(i)}{\mathcal{N}_{spcl}(i)} \right) \tag{12}$$

$$\mathcal{L}^{spcl} = \sum_{i=1}^{N} \mathcal{L}_i^{spcl} \tag{13}$$

Method

Curriculum Learning

Let the total size of training set $\mathcal{D}_{train}$ as L,

$$\mathbf{C}_k = \frac{1}{|\{z_j|\forall j, y_j = k\}|} \sum_{j=1}^{L} z_j \cdot \mathbb{I}(y_j = k) \quad (14)$$

$$\mathcal{DIF}(i) = \frac{\mathrm{dis}(z_i, \mathbf{C}_{y_i})}{\sum_{j=1}^{|\mathcal{E}|} \mathrm{dis}(z_i, \mathbf{C}_j)} \quad (15)$$

$$a_1 = 1 - k/R \text{ and } a_L = k/R.$$

hard. Let R as the number of training epochs, to train the model at k-th epoch, we first generate a arithmetic progression a with a length of L, where $a_1 = 1 - k/R$ and $a_L = k/R$. Then we initialize a Bernoulli distribution with a and draw a binary random array R_B from it. We use B to draw a subset $\mathcal{D}_{sub-k}$ from training set for the current epoch, where $\mathcal{D}_{sub-k} \equiv \{x_i \in \mathcal{D}_{train} | R_{Bi} = 1\}$. Ob-

$$p_m^{ic} = \frac{\mathcal{G}(z_i, \mathbf{C}_c)}{\sum_{k=1}^{|\mathcal{E}|} \mathcal{G}(z_i, \mathbf{C}_k)} \quad (16)$$

$$p_l^i = \mathrm{W} \cdot z_i + b \quad (17)$$

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{\mathcal{E}} y_{ic} \cdot \log p_l^{ic} \quad (18)$$

Experiment

	MELD	IEMOCAP	EmoryNLP
No.Dials	1,432	151	827
Train	1,038	100	659
Dev	114	20	89
Test	280	31	79
No.Uttrs	13,708	7,333	9,489
Train	9,989	4,810	7,551
Dev	1,109	1,000	954
Test	2,610	1,523	984
No.CLS	7	6	7

Table 2: Statistics of the three datasets.

Experiment

Models	IEMOCAP	MELD	EmoryNLP
COSMIC(Ghosal et al., 2020)	65.28	65.21	38.11
DialogueCRN (Hu et al., 2021)	66.46	63.42	38.91
DAG-ERC (Shen et al., 2021)	68.03	63.65	39.02
TODKAT (Zhu et al., 2021)	61.33	65.47	38.69
Cog-BART (Li et al., 2021)	66.18	64.81	39.04
TUCORE-GCN_RoBERTa(Lee and Choi, 2021)	-	65.36	39.24
SGED + DAG-ERC(Bao et al., 2022)	68.53	65.46	40.24
EmotonFlow-Large (Song et al., 2022)	-	66.50	-
CoMPM (Lee and Lee, 2021)	69.46	66.52	38.93
SPCL-CL-ERC(Ours)	**69.74**	**67.25**	**40.94**

Table 1: Performance comparisons on three datasets.

Experiment

	IEMOCAP	MELD	EmoryNLP
CE	68.35	65.33	38.72
+ CL	67.40	65.63	39.00
SupCon	68.13	65.67	39.20
+ CL	68.64	66.15	39.49
SPCL	69.03	66.56	40.14
+ CL	69.74	67.25	40.94

Table 3: Results of ablation study. Here, CE means Cross-entropy loss, SupCon is the vanilla supervised contrastive learning loss and SPCL is our proposed supervised prototypical contrastive learning loss. CL is our proposed curriculum strategy.

Experiment

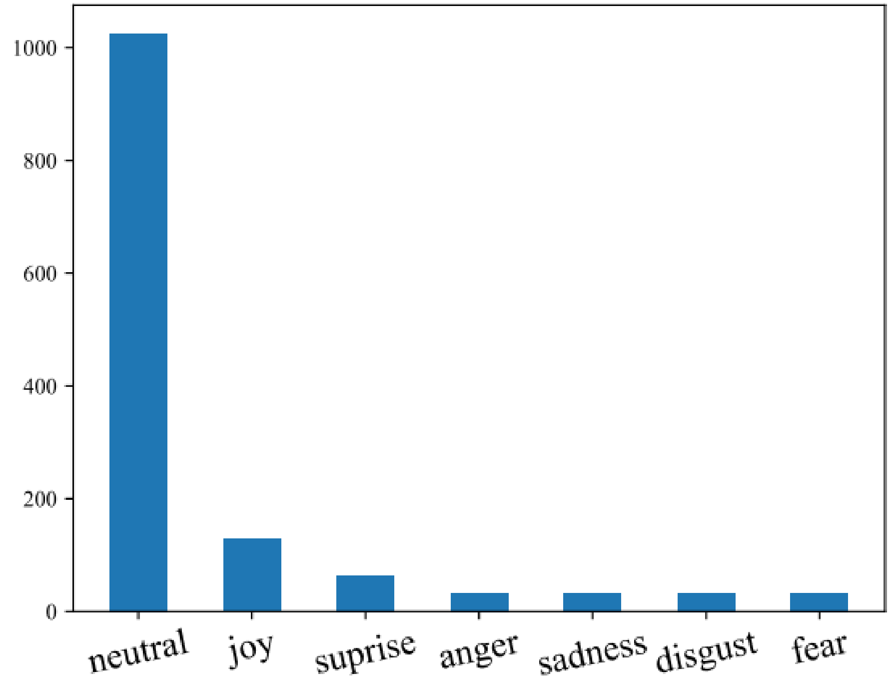


Figure 4: Emotion distribution of the extreme class-imbalanced training set. We construct it from MELD training set.

	4	8	16	32
SupCon	53.14	57.36	58.50	60.09
SPCL	57.27	58.85	59.47	61.38

Table 4: Results of different loss functions and different batch sizes trained on the imbalanced training set.

	4	8	16	32
SupCon	62.50	65.01	67.04	68.13
SPCL	68.21	68.41	68.48	69.03

Table 5: Results of SupCon and SPCL with different batch sizes on IEMOCAP dataset.

Experiment

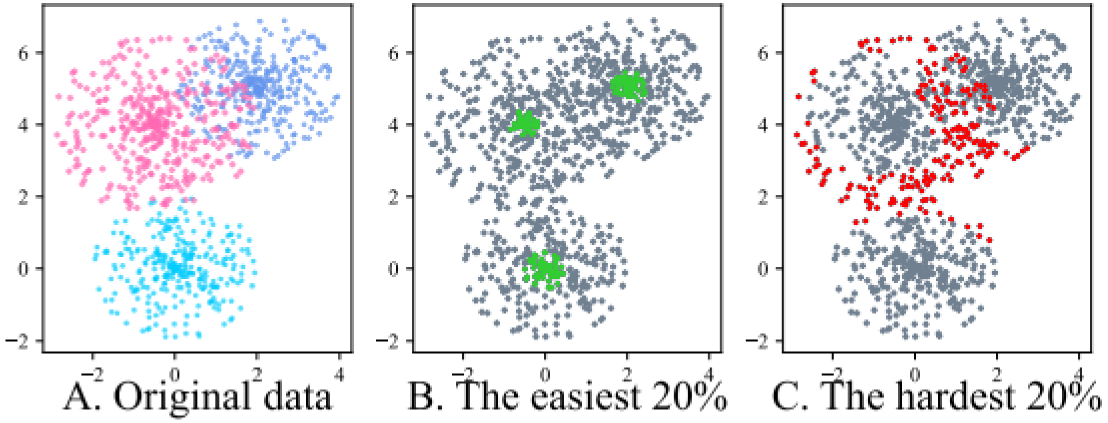


Figure 5: Visualizations of how the difficulty measure function $\mathcal{DIF}$ in Eq.(15) ranks the data.

Experiment

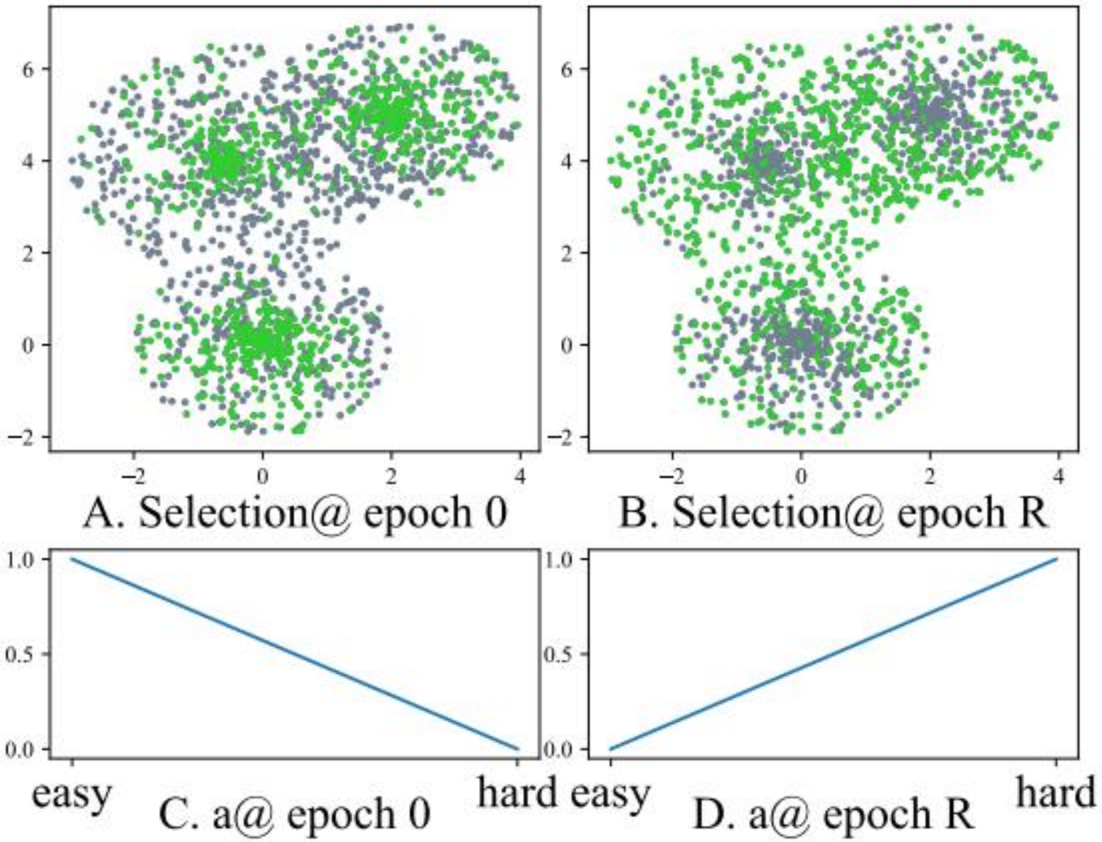


Figure 6: The sampling-based curriculum strategy.

Thank you!