Emotion Detection from Text

2022/11/18

DialogueCRN: Contextual Reasoning Network for Emotion Recognition in Conversations

DialogueCRN

- 近年有关ERC(Emotion Recognition in Conversations)工作：通过深度学习模型感知情境级或说话人级上下文
- **情绪认知理论**：认知因素对情绪状态的形成有重要影响；认知因素通过反复执行大脑中的直觉检索过程和有意识地推理过程来捕捉
- **感知阶段**：LSTM网络捕捉情境级、说话人级上下文
- **认知阶段**：多回合推理模块，执行两个过程
  - 直观检索过程：利用注意力机制检索静态全局记忆，匹配相关线索
  - 有意识推理过程：LSTM网络学习内在逻辑顺序，通过保留和更新动态记忆来整合上下文线索
- 根据情境级和说话人级的上下文线索，情感分类器预测情感标签

Perception Phase

Conversational Context Representation:

情境级别： $\mathbf{c}_i^s, \mathbf{h}_i^s = \overleftrightarrow{LSTM}^s(\mathbf{u}_i, \mathbf{h}_{i-1}^s),$

说话人级别：$\mathbf{c}_i^v, \mathbf{h}_{\lambda,j}^v = \overleftrightarrow{LSTM}^v(\mathbf{u}_i, \mathbf{h}_{\lambda,j-1}^v), j \in [1, |U_\lambda|],$

Global Memory Representation:

$$\mathbf{g}_i^s = \mathbf{W}_g^s \mathbf{c}_i^s + \mathbf{b}_g^s,$$

$$\mathbf{g}_i^v = \mathbf{W}_g^v \mathbf{c}_i^v + \mathbf{b}_g^v,$$

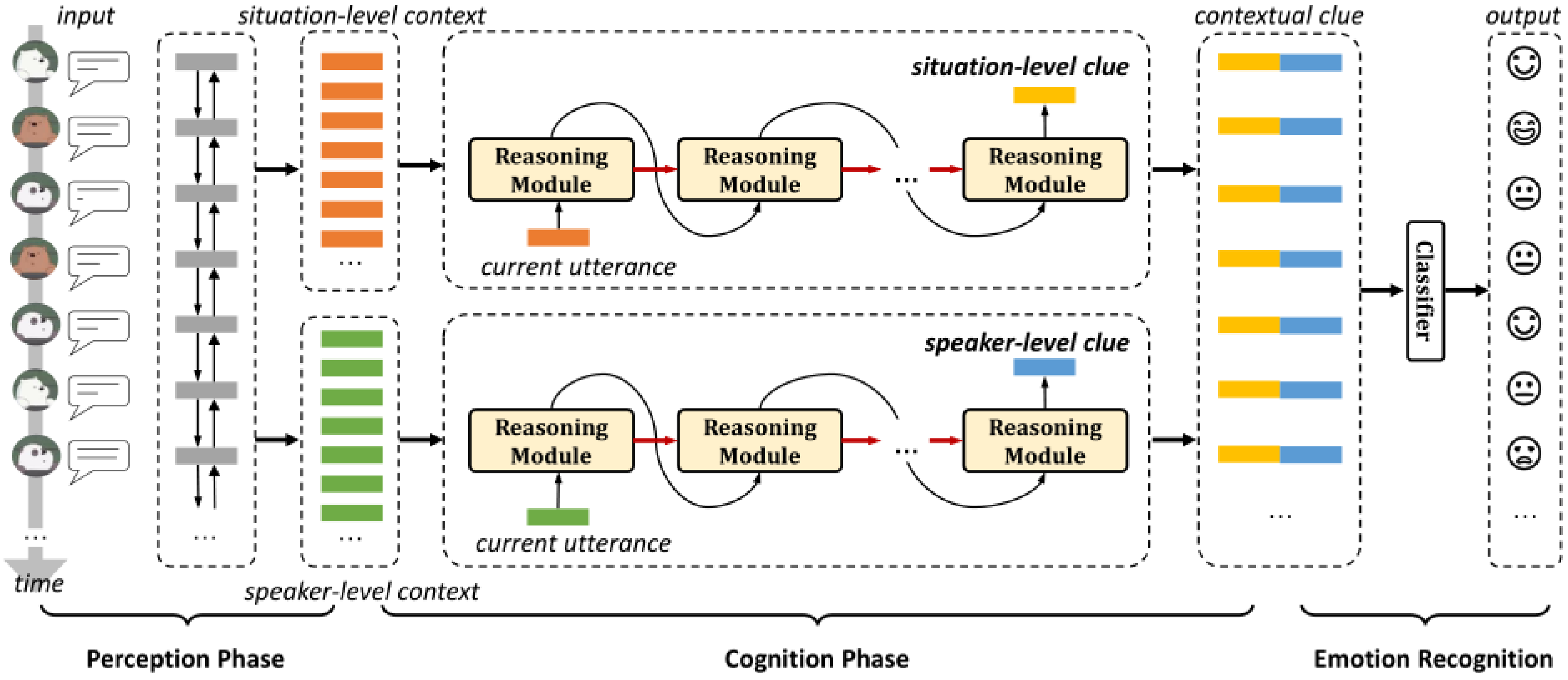


Figure 1: The architecture of the proposed model DialogueCRN.

Cognition Phase

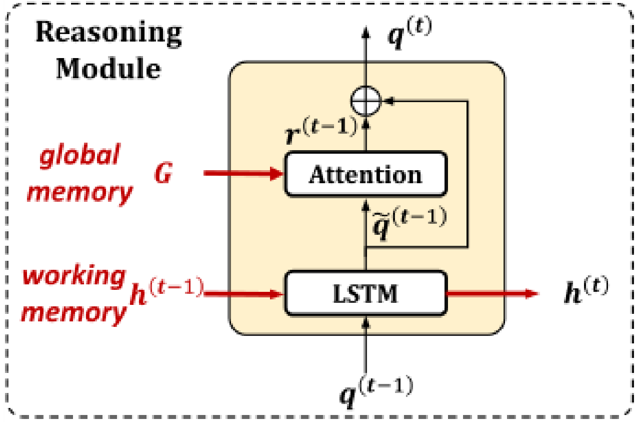


Figure 2: The detailed structure of reasoning module.

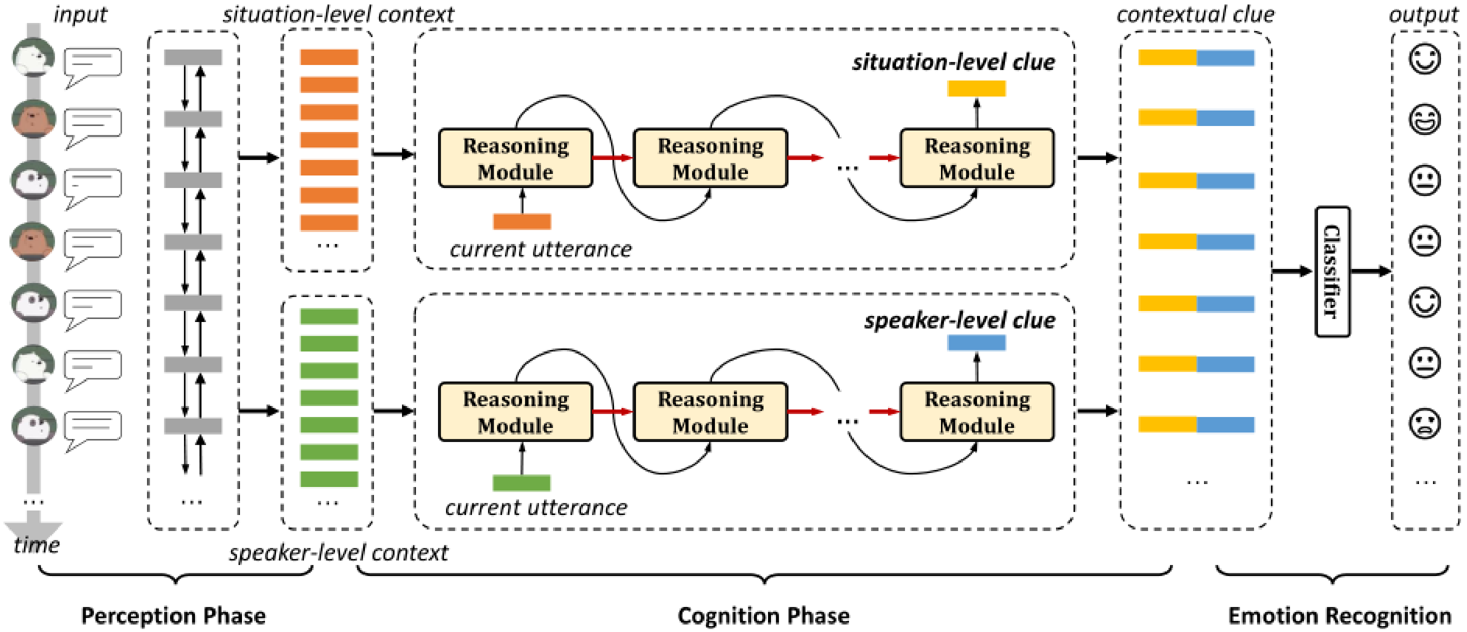


Figure 1: The architecture of the proposed model DialogueCRN.

Reasoning process:

$$\tilde{\mathbf{q}}_i^{(t-1)}, \mathbf{h}_i^{(t)} = \overrightarrow{LSTM}(\mathbf{q}_i^{(t-1)}, \mathbf{h}_i^{(t-1)}),$$

Retrieving process:

$$\mathbf{e}_{ij}^{(t-1)} = f(\mathbf{g}_j, \tilde{\mathbf{q}}_i^{(t-1)}),$$

$$\alpha_{ij}^{(t-1)} = \frac{\exp(\mathbf{e}_{ij}^{(t-1)})}{\sum_{j=1}^{N} \exp(\mathbf{e}_{ij}^{(t-1)})},$$

$$\mathbf{r}_i^{(t-1)} = \sum_{j=1}^{N} \alpha_{ij}^{(t-1)} \mathbf{g}_j,$$

$$\mathbf{q}_i^{(t)} = [\tilde{\mathbf{q}}_i^{(t-1)}; \mathbf{r}_i^{(t-1)}].$$

To sum up:

$$\mathbf{q}_i^s = Cognition^s(\mathbf{c}_i^s, \mathbf{G}^s; T^s),$$

$$\mathbf{q}_i^v = Cognition^v(\mathbf{c}_i^v, \mathbf{G}^v; T^v),$$

$$\mathbf{o}_i = [\mathbf{q}_i^s; \mathbf{q}_i^v].$$

Experimental setups

Dataset	# Dialogues train	# Dialogues val	# Dialogues test	# Utterances train	# Utterances val	# Utterances test	Avg. Length	# Classes
IEMOCAP	120		31	5,810		1,623	50	6
SEMAINE	63		32	4,368		1,430	72	4*
MELD	1,039	114	280	9,989	1,109	2,610	10	7

* refers to the number of real valued attributes.

Table 1: The statistics of three datasets.

IEMOCAP			
Methods	*Acc.*	*Weighted-F1*	*Macro-F1*
TextCNN	49.35	49.21	48.13
Memnet	55.70	53.10	55.40
bc-LSTM+Att	56.32	56.19	54.84
CMN	56.56	56.13	54.30
ICON	59.09	58.54	56.52
DialogueRNN	63.03	62.50	60.66
DialogueGCN	64.02	63.65	63.43
DialogueCRN	**66.05**	**66.20**	**66.38**
Improve	**3.2%**	**4.0%**	**4.7%**

Table 2: Experimental results on the *IEMOCAP*

SEMAINE				
Methods	*MAE*			
	Valence	*Arousal*	*Expectancy*	*Power*
TextCNN	0.545	0.542	0.605	8.71
Memnet	0.202	0.211	0.216	8.97
bc-LSTM+Att	0.189	0.213	0.190	8.67
CMN	0.192	0.213	0.195	8.74
ICON	0.180	0.190	0.180	8.45
DialogueRNN	0.175	0.171	0.181	8.66
DialogueGCN	0.176	0.210	0.193	8.65
DialogueCRN	**0.173**	**0.152**	**0.175**	**8.20**
Improve	**1.1%**	**11.1%**	**2.8%**	**2.9%**

Table 3: Experimental results on the *SEMAINE* dataset.

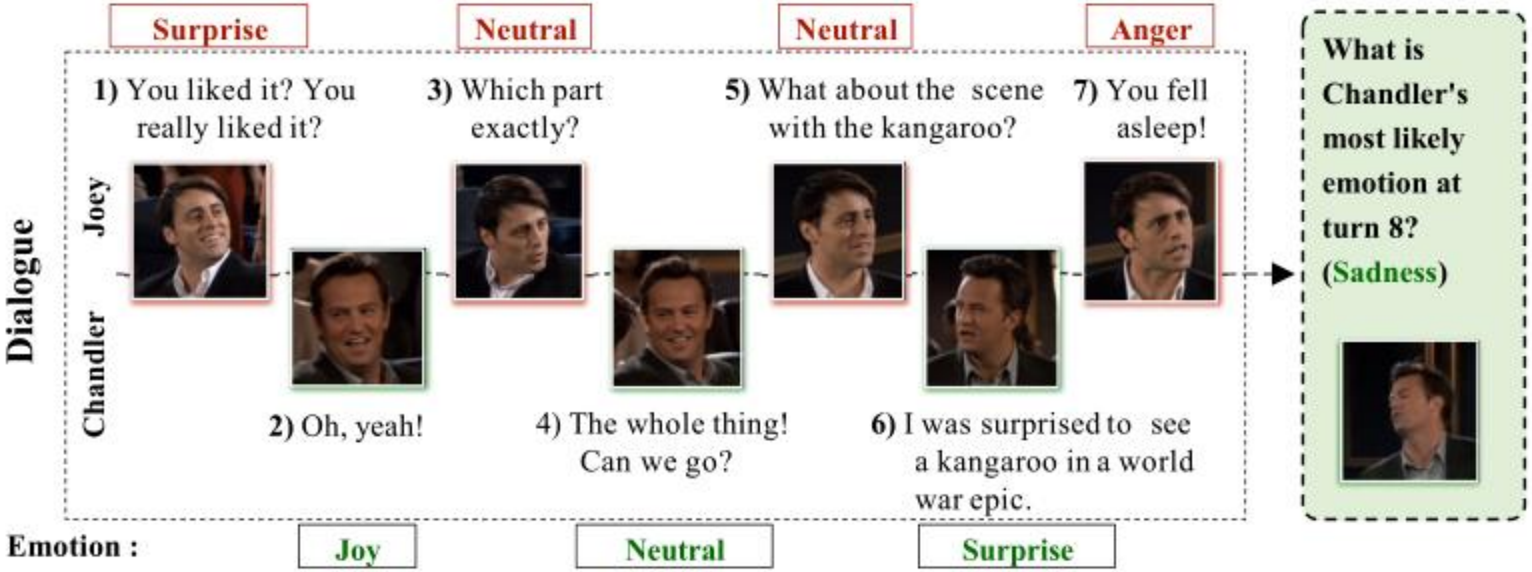
MELD			
Methods	*Acc.*	*Weighted-F1*	*Macro-F1*
TextCNN	59.69	56.83	33.80
bc-LSTM+Att	57.50	55.90	34.84
CMN	-	54.50	-
ICON	-	54.60	-
DialogueRNN	59.54	56.39	32.93
DialogueGCN	59.46	56.77	34.05
DialogueCRN	**60.73**	**58.39**	**35.51**
Improve	**2.0%**	**2.9%**	**1.9%**

Table 4: Experimental results on the *MELD* dataset.

Emotion Inference in Multi-Turn Conversations with Addressee-Aware Module and Ensemble Strategy

Overview

- 研究多回合会话中的情绪推断任务
- 对会话历史中参与者间的情绪状态传播进行建模
- Addressee-aware模块：建模情感特征，预测情感趋势
- Ensemble strategy：进一步提高性能



Task Definition

- 数据表示：D = $[(U_1, p_1^w), (U_2, p_2^w), \cdots, (U_m, p_m^w), p_{m+1}^a]$
- 预测任务：$E_{m+1}^a \sim P(E_{m+1}^a | (U_1, p_1^w), \cdots, (U_m, p_m^w), p_{m+1}^a)$.

Feature Extraction

$$u_1, u_2, \cdots, u_m = \mathrm{CNN/RoBERTa}(U_1, U_2, \cdots, U_m),$$

Addressee-Aware Module

- 持续性（persistence）：参与者可能持续受自己情绪的影响，并在一段时间内保持现有的情绪状态
- 传染性（contagiousness）：参与者的情绪状态的互动的、有影响力的、相互传染的
- Addressee-Aware Module使用基于序列、基于图形的模型同时对上述两种情绪建模

Addressee-Aware Module

- Sequence-based Model
  - Persistence：LSTM-store打开输入门，尽可能多地将Ut存入内部状态
  - Contagiousness：
    - 若是高度感染性、很可能影响情绪：LSTM-affect打开遗忘门，忘记过去状态，用另一参与者的Ut更新当前状态
    - 否则， LSTM-affect关闭输入门，保留历史状态，转换为t时刻的内部状态

$$
\begin{aligned}
(h_t^a, c_t^a) =& \lambda_t^a \cdot LSTM_{store}(u_t, (h_{t-1}^a, c_{t-1}^a)) \\
&+ (1 - \lambda_t^a) \cdot LSTM_{affect}(u_t, (h_{t-1}^a, c_{t-1}^a)),
\end{aligned}
$$

$$
\lambda_t^a = \begin{cases} 1, if\ p_t^w = p_{m+1}^a \\ 0, if\ p_t^w \neq p_{m+1}^a \end{cases},
$$

$$
es_{m+1}^a = \mathrm{softmax}\left(\mathbf{W}_c^T \left(\mathrm{ReLU}(\mathbf{W}_s^T h_m^a)\right) + b\right),
$$

Addressee-Aware Module

- Graph-based Model
  - 为每个对话构造有向图G = （g，e，α）
  - 节点表示： $g_t = (\mathbf{W}_l^T u_t + b),$
  - 权重计算：ATT-store，ATT-affect

$$\alpha_{m,t}^a = \mathrm{softmax}(\lambda_t^a \cdot ATT_{store}(g_m, g_t) \\ +(1-\lambda_t^a) \cdot ATT_{affect}(g_m, g_t)),$$

$$\lambda_t^a = \begin{cases} 1, if\ p_t^w = p_{m+1}^a \\ 0, if\ p_t^w \neq p_{m+1}^a \end{cases},$$

$$ATT(g_m, g_t) = \mathbf{W}_a^T \left(\mathrm{ReLU}\left(\mathbf{W}_f^T [g_m \| g_t]\right)\right),$$

$${g_m}' = \sum\nolimits_{g_t \in \mathcal{H}_{g_m}} \alpha_{m,t}^a \cdot g_t,$$

$$eg_{m+1}^a = \mathrm{softmax}\left(\mathbf{W}_c^T \left(\mathrm{ReLU}(\mathbf{W}_g^T g_m')\right) + b\right)$$

Ensemble Strategy

- DialogInfer-(S+G)： $ei_{m+1}^{a} = \mathrm{softmax}\left(\mathbf{W}_c^T\left(\mathrm{ReLU}(\mathbf{W}_i^T(h_m^a + g_m^{'}))\right) + b\right)$
- DialogInfer-Ensemble：用不同seed分别训练DialogInfer-S、DialogInfer-G、DialogInfer-(S+G)5次，训练融合分类器：

$$ef_{m+1}^{a} = \mathrm{softmax}(\mathrm{ReLU}(\mathbf{W}_f^T([es_{m+1}^{a}{}^{1}||\cdots||es_{m+1}^{a}{}^{5}|| \\ eg_{m+1}^{a}{}^{1}||\cdots||eg_{m+1}^{a}{}^{5}||ei_{m+1}^{a}{}^{1}||\cdots||ei_{m+1}^{a}{}^{5}]) + b)),$$

- 最终分类结果从四种类型的模型输出概率分布中采样：

$$E_{m+1}^{a} \sim P(E_{m+1}^{a}|(U_1, p_1^w), \cdots, (U_m, p_m^w), p_{m+1}^{a}) \\ = (es_{m+1}^{a}/eg_{m+1}^{a}/ei_{m+1}^{a}/ef_{m+1}^{a}).$$

Experiments

	Methods	IEMOCAP	MELD	EmoryNLP
GloVe-based	CNN (2014)	44.09	36.31	20.97
	sc-LSTM (2017)	56.22	36.06	19.75
	DialogueRNN (2019)	58.12	36.93	20.37
	DialogueGCN (2019)	56.48	36.98	19.59
	DialogInfer-S	60.45	38.09	21.08
	DialogInfer-G	59.48	36.62	20.22
	DialogInfer-(S+G)	60.74	38.46	**21.69**
	DialogInfer-Ensemble	**65.31***	**38.48***	20.95
RoBERTa-based	RoBERTa Large (2019)	43.24	36.99	20.46
	sc-LSTM (2017)	58.81	37.71	22.26
	DialogueRNN (2019)	59.53	38.70	21.98
	COSMIC (2020)	61.50	39.49	21.60
	DialogInfer-S	63.63	40.32	23.09
	DialogInfer-G	59.94	38.06	22.81
	DialogInfer-(S+G)	64.70	**40.67**	22.63
	DialogInfer-Ensemble	**66.39***	39.41	**24.09***

Table 1: Performance on three datasets. The weighted

	Methods	IEMOCAP	MELD	EmoryNLP
GloVe-based	**DialogInfer-S**	**60.45**	**38.09**	**21.08**
	w/o addressee-aware	57.04	36.30	18.94
	DialogInfer-G	**59.48**	**36.62**	**20.22**
	w/o addressee-aware	56.42	35.21	20.10
	DialogInfer-(S+G)	**60.74**	**38.46**	**21.69**
	w/o addressee-aware	58.79	37.33	19.70
	DialogInfer-Ensemble	**65.31**	**38.48**	**20.95**
	w/o addressee-aware	58.72	36.75	20.73
RoBERTa-based	**DialogInfer-S**	**63.63**	**40.32**	**23.09**
	w/o addressee-aware	59.23	38.03	22.52
	DialogInfer-G	**59.94**	**38.06**	**22.81**
	w/o addressee-aware	56.43	37.17	21.03
	DialogInfer-(S+G)	**64.70**	**40.67**	**22.63**
	w/o addressee-aware	59.39	38.81	22.16
	DialogInfer-Ensemble	**66.39**	**39.41**	**24.09**
	w/o addressee-aware	61.16	37.85	21.79

Table 2: Ablation analysis on three datasets.

Knowledge-Enriched Transformer for Emotion Detection in Textual Conversations

Contributions

- 从外部知识库和情感词典中获得动态、上下文感知的情感常识
- 分层自注意力机制：对对话的分层结构进行建模

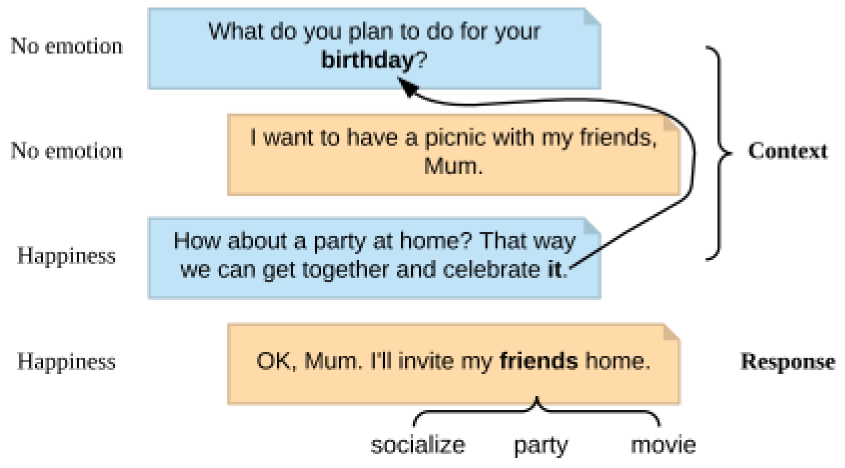


Figure 1: An example conversation with annotated labels from the DailyDialog dataset (Li et al., 2017). By referring to the context, "it" in the third utterance is linked to "birthday" in the first utterance. By leveraging an external knowledge base, the meaning of "friends" in the forth utterance is enriched by associated knowledge entities, namely "socialize", "party", and "movie". Thus, the implicit "happiness" emotion in the fourth utterance can be inferred more easily via its enriched meaning.

Methodology

- Knowledge Retrieval
  - ConceptNet：<concept1, relation, concept2>
  - NRC_VAD：英文单词与其VAD分数的列表（Valence, Arousal, Dominance）
  - 去除停用词与不在NRC_VAD中的词： (c1,s1,VAD(c1)), (c2,s2,VAD(c2))…

- Embedding Layer

$$\mathbf{t} = Embed(t) + Pos(t).$$

Methodology

- Dynamic Context-Aware Affective Graph Attention
  - 概念表示：

$$\alpha_k = softmax(w_k), \qquad \mathbf{c}(t) = \sum_{k=1}^{|g(t)|} \alpha_k * \mathbf{c_k},$$

  - 假设重要的概念是那些与会话上下文相关且具有强烈情感强度的概念
    - Relatedness：衡量Ck与对话上下文的关联程度

$$\mathbf{CR}(X^i) = avg(\mathbf{SR}(X^i_{j-M}), ..., \mathbf{SR}(X^i_j)),$$

$$rel_k = min\text{-}max(s_k) * abs(\cos(\mathbf{CR}(X^i), \mathbf{c_k})),$$

    - Affectiveness：度量Ck的情感强度

$$aff_k = min\text{-}max(||[V(c_k) - 1/2, A(c_k)/2]||_2),$$

  - 最终表示：

$$w_k = \lambda_k * rel_k + (1 - \lambda_k) * aff_k,$$
$$\hat{\mathbf{t}} = \mathbf{W}[\mathbf{t}; \mathbf{c}(t)],$$

Methodology

- Hierarchical Self-Attention
  - Utterance-level：话语级自注意力层计算每个话语表示

$$\hat{\mathbf{X}}'^{i}_{n} = FF(L^{'}(MH(L(\hat{\mathbf{X}}^{i}_{n}), L(\hat{\mathbf{X}}^{i}_{n}), L(\hat{\mathbf{X}}^{i}_{n})))),$$

$$MH(Q, K, V) = softmax(\frac{QK^{T}}{\sqrt{d_s}})V,$$

$$FF(x) = max(0, xW_1 + b_1)W_2 + b_2,$$

  - Context-level：上下文级自注意力层从M个话语表示中计算上下文表示

$$\mathbf{C}^{i} = FF(L^{'}(MH(L(\hat{\mathbf{X}}^{i}), L(\hat{\mathbf{X}}^{i}), L(\hat{\mathbf{X}}^{i})))),$$

Experiments

Dataset	Domain	#Conv. (Train/Val/Test)	#Utter. (Train/Val/Test)	#Classes	Evaluation
EC	Tweet	30160/2755/5509	90480/8265/16527	4	Micro-F1
DailyDialog	Daily Communication	11118/1000/1000	87170/8069/7740	7	Micro-F1
MELD	TV Show Scripts	1038/114/280	9989/1109/2610	7	Weighted-F1
EmoryNLP	TV Show Scripts	659/89/79	7551/954/984	7	Weighted-F1
IEMOCAP	Emotional Dialogues	100/20/31	4810/1000/1523	6	Weighted-F1

Table 1: Dataset descriptions.

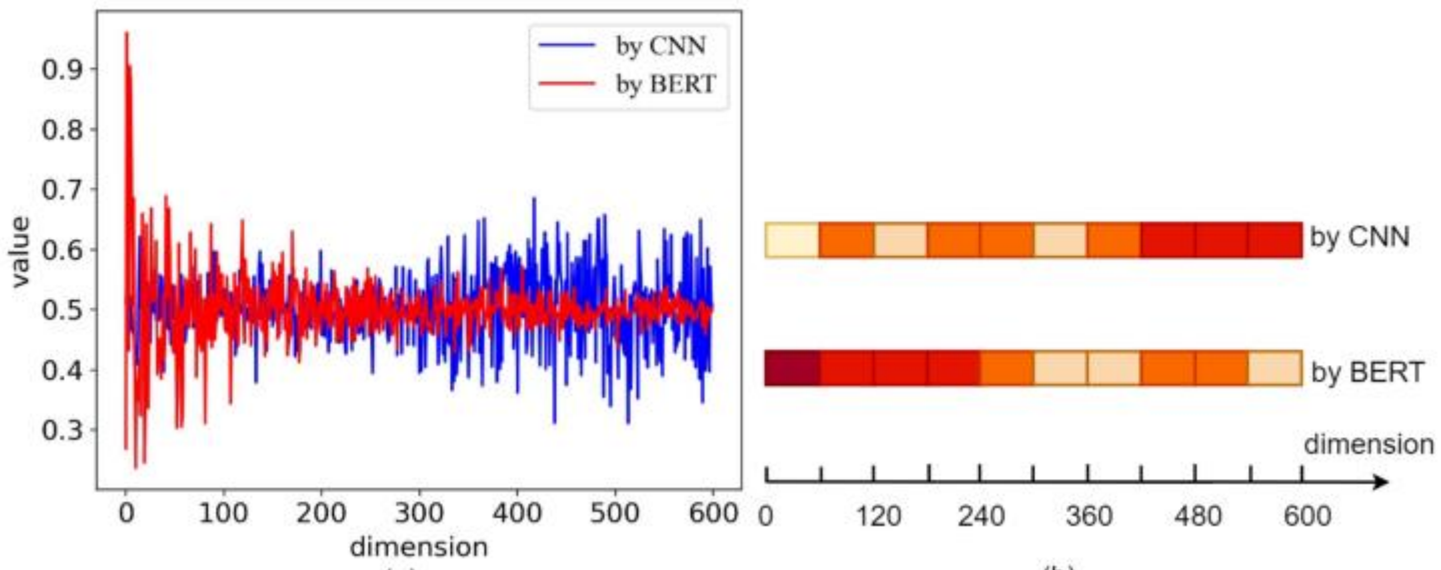
Model	EC	DailyDialog	MELD	EmoryNLP	IEMOCAP
cLSTM	0.6913	0.4990	0.4972	0.2601	0.3484
CNN (Kim, 2014)	0.7056	0.4934	0.5586	0.3259	0.5218
CNN+cLSTM (Poria et al., 2017)	0.7262	0.5024	0.5687	0.3289	0.5587
BERT_BASE (Devlin et al., 2018)	0.6946	0.5312	0.5621	0.3315	0.6119
DialogueRNN (Majumder et al., 2019)	0.7405	0.5065	0.5627	0.3170	**0.6121**
KET_SingleSelfAttn (ours)	0.7285	0.5192	0.5624	0.3251	0.5810
KET_StdAttn (ours)	**0.7413**	0.5254	0.5682	0.3353	0.5861
KET (ours)	0.7348	**0.5337**	**0.5818**	**0.3439**	0.5956

Table 2: Performance comparisons on the five test sets. Best values are highlighted in bold.

SEOVER: Sentence-level Emotion Orientation Vector based Conversation Emotion Recognition Model

Motivation

- 对话情感识别：说话人识别、对话关系建模
- CNN编码：表达话语的语法和语义能力欠佳
- BERT编码：无法完全提取句子之间情绪倾向的相关性



(a) (b)

SEOV & SEOVER

- SEOV：句子级别情感定向向量
  - Size：句子编码向量的情绪强度
  - Direction：模拟向量间情感趋势

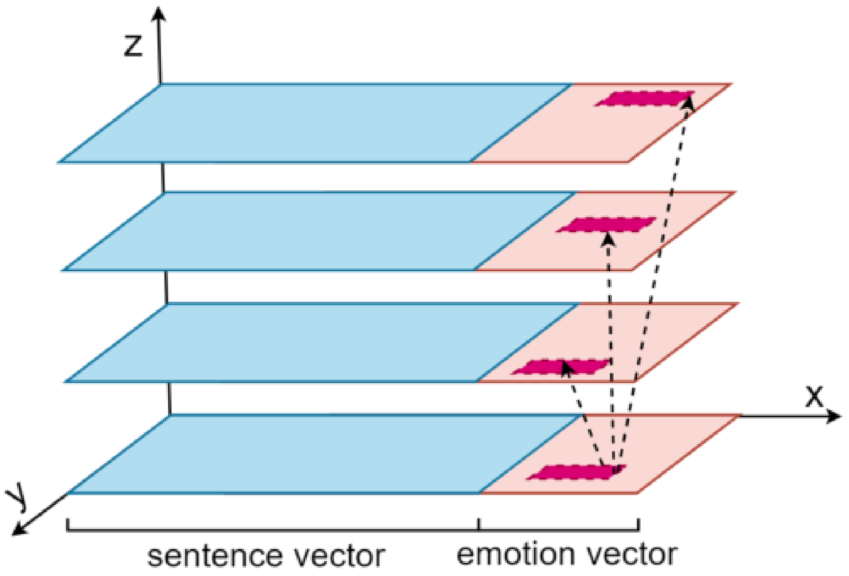
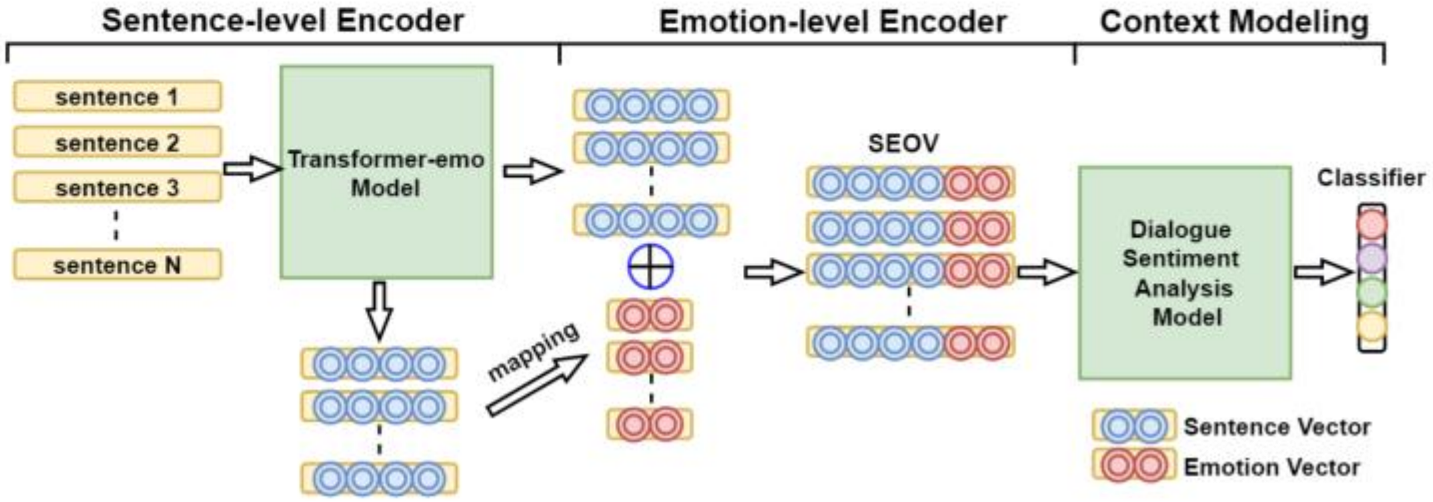


Fig. 2. The schematic diagram of SEOV, where the x-axis represents the dimension of the vector, y-axis represents the value of each dimension of the vector, and z-axis represents different sentence-level emotion orientation vectors. There is a correlation between different emotion vectors, indicating the "direction" of the SEOV.

- SEOVER：对话情感识别模型
  - Transformer-emo提取SEOV
  - SEOV和情感分析模型获得上下文语义信息
  - 微调得情感分类结果

Methodology



- Sentence-level Encoder： $Q = Transformer - emo(s_1, s_2, s_3 \dots s_n)$
- Emotion-level Encoder： $q* = q[w_1, w_2, \dots w_{k^*}]$

$$e = q \oplus q^*$$

- Context Modeling：将SEOV组合成对话列表，再输入DSAM分类

Experiments

Table 1. Experimental results(F1 score) on the IEMOCAP dataset and MELD dataset. SEOVER-RNN represents the result of using DialogueRNN as the fine-tuning model, SEOVER-GCN represents the result of using DialogueGCN as the fine-tuning model, and SEOVER-LSTM represents the result of using bc-LSTM as the fine-tuning model.

Model	IEMOCAP							MELD
	Happy	Sad	Neutral	Angry	Excited	Frustrated	Average	Average
DialogueRNN	33.18	78.80	59.21	65.28	71.86	58.91	62.75	55.90
DialogueGCN	42.75	84.54	63.54	64.19	63.08	**66.99**	64.18	-
DialogXL	-	-	-	-	-	-	65.95	62.41
TRMSM	50.22	75.82	64.15	60.97	72.70	63.45	65.74	62.36
bc-LSTM	43.40	69.82	55.84	61.80	59.33	60.20	59.19	55.90
BERT	-	-	-	-	-	-	54.01	60.34
SEOVER-RNN	69.47	83.57	**66.67**	67.46	**82.46**	55.36	69.86	**65.66**
SEOVER-GCN	53.85	80.40	58.77	62.09	79.22	56.87	65.29	-
SEOVER-LSTM	**70.38**	**85.65**	65.40	**69.45**	80.98	64.96	**72.07**	63.82

Table 2. Results of ablation experiments, where accuracy and F1 score are both weighted results.

Model	Accuracy	F1 score
DialogueRNN	56.10	55.90
DialogueRNN-BERT	58.59	59.40
SEOVER-RNN	**65.33**	**65.66**

分享结束，谢谢大家！

2022/11/18