TSAM: A Two-Stream Attention Model for Causal Emotion Entailment

Duzhen Zhang* , Zhen Yang, Fandong Meng, Xiuyi Chen, Jie Zhou
Pattern Recognition Center, WeChat AI, Tencent Inc, Beijing, China
{bladedancer957,hugheren.chan}@gmail.com
{zieenyang,fandongmeng,withtomzhou}@tencent.com

Code：https://github.com/BladeDancer957/TSAM
Data：https://github.com/declare-lab/RECCON/tree/main/data

2023.4.15 • ChongQing

—— COLING 2022

Reported by Renhui Luo

1.Introduction

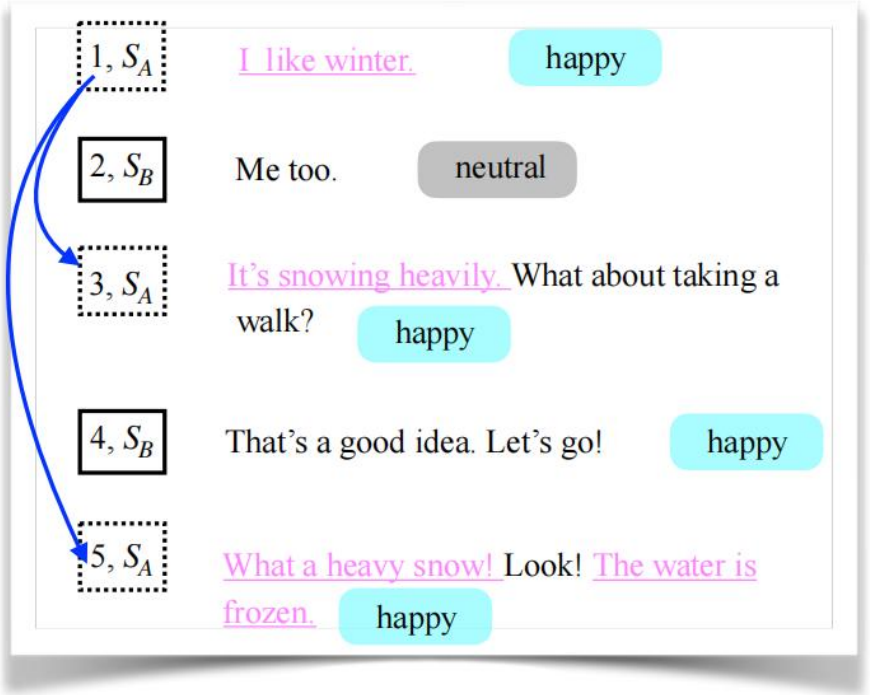
2.Overview

3.Methods

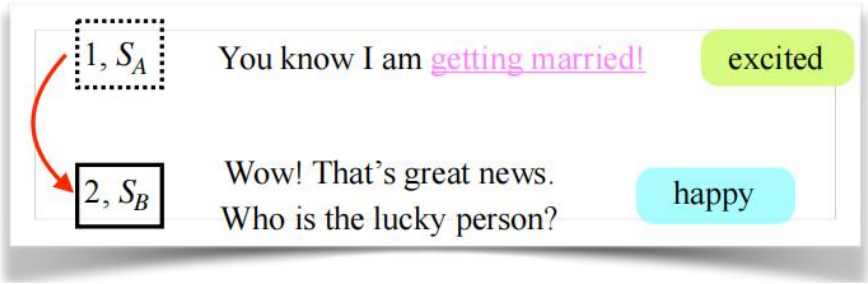
4.Experiments

Introduction

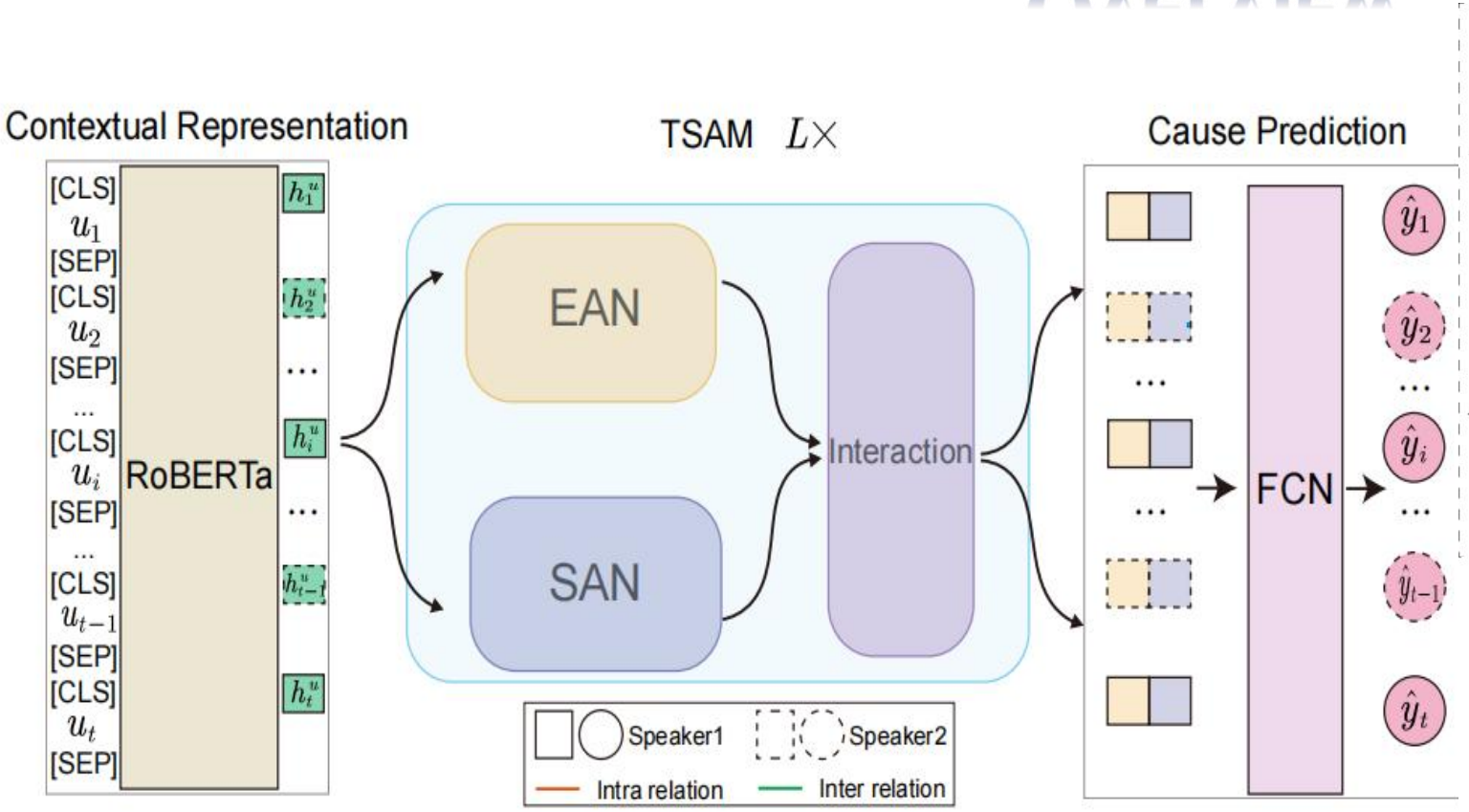
a)

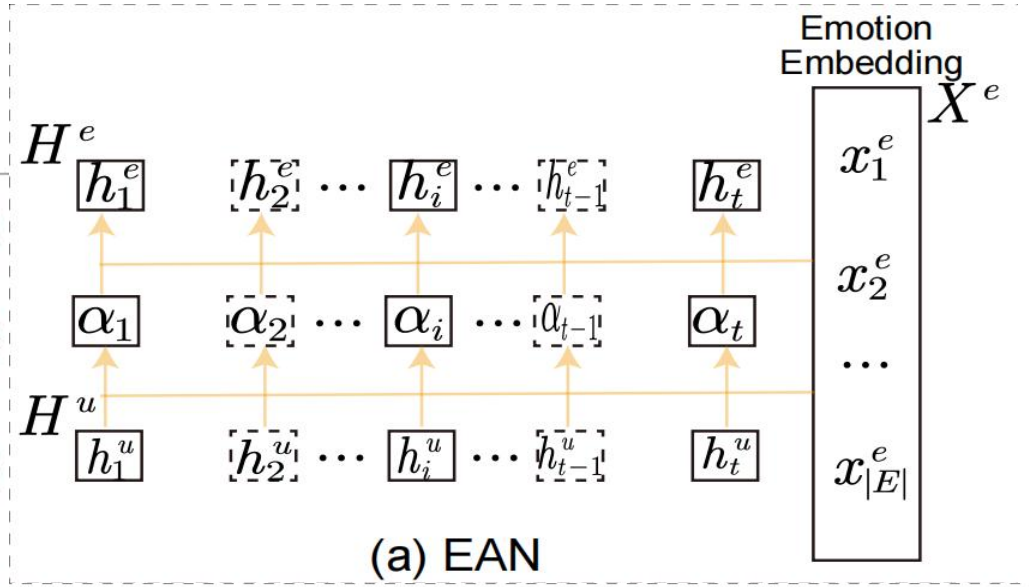


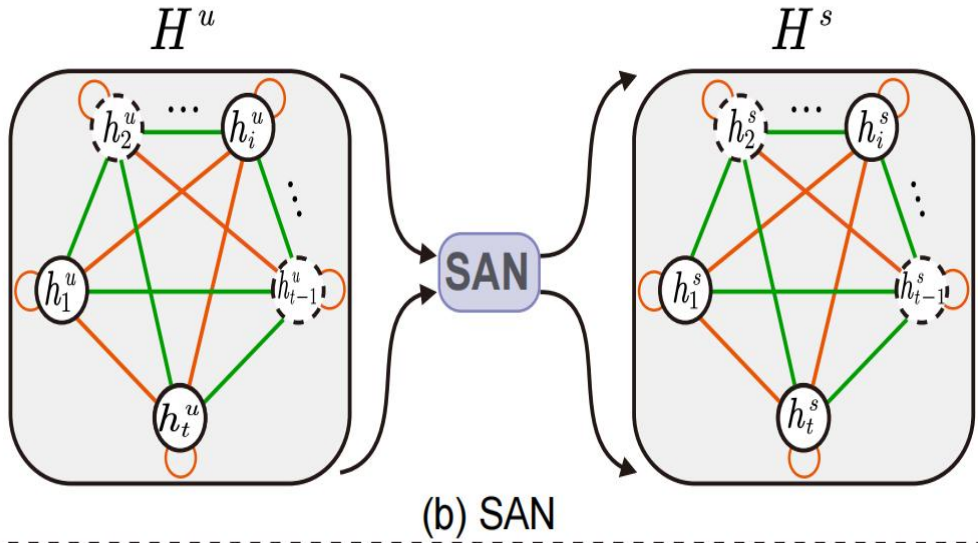
b)



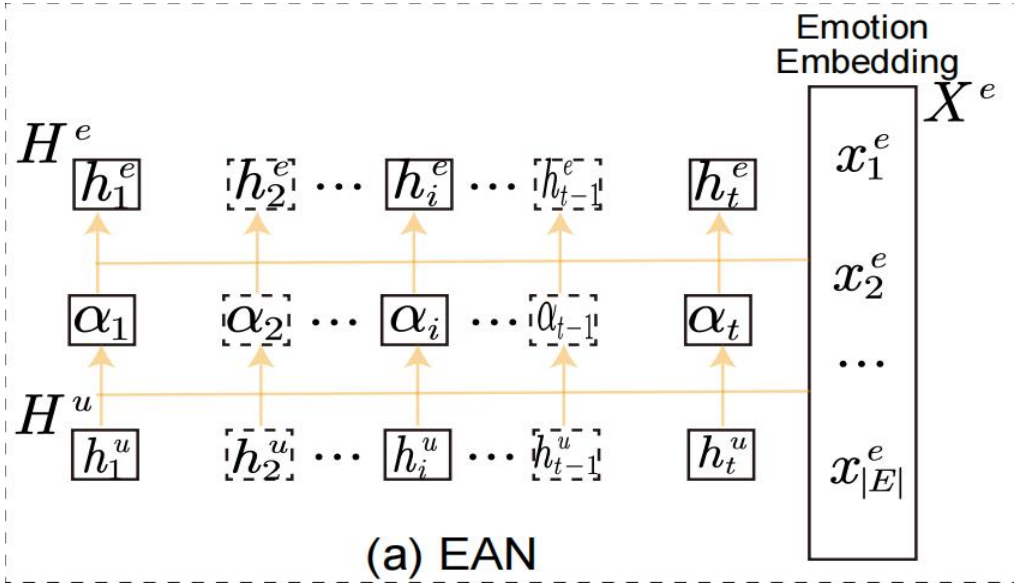
Overview


Contextual Representation
RoBERTa
TSAM $L\times$
EAN
SAN
Interaction
Speaker1
Speaker2
Intra relation
Inter relation
Cause Prediction
FCN


Emotion Embedding
(a) EAN


SAN
(b) SAN

Method



(a) EAN

$$\boldsymbol{x}_k^e = \text{Embed}(e_k) \in \mathbb{R}^{d_h} \quad (1)$$

$$\boldsymbol{H}^e = attention(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \boldsymbol{\alpha}\boldsymbol{V} \quad (2)$$

$$\boldsymbol{\alpha} = softmax(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d_h}}) \quad (3)$$

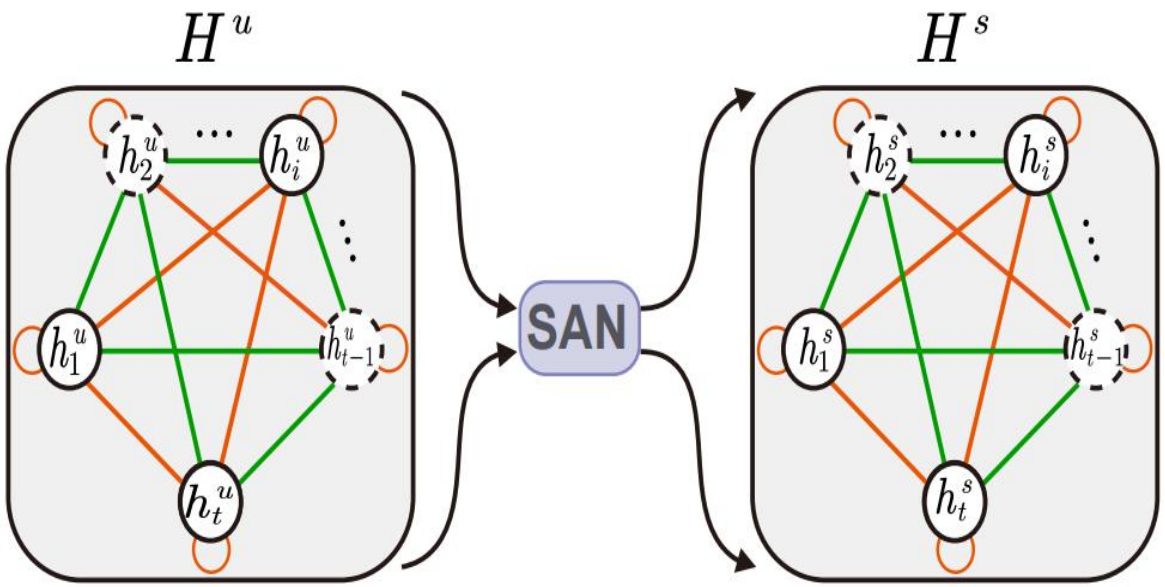
$$\boldsymbol{H}^e = concat(head_1, ..., head_m) \quad (4)$$

$$head_j = attention(\boldsymbol{Q}\boldsymbol{W}_j^Q, \boldsymbol{K}\boldsymbol{W}_j^K, \boldsymbol{V}\boldsymbol{W}_j^V) \quad (5)$$

$$\boldsymbol{H}^e = \tilde{\boldsymbol{X}}^e \quad (6)$$

where $\boldsymbol{Q} = \boldsymbol{H}^u, \boldsymbol{K} = \boldsymbol{V} = \boldsymbol{X}^e,$

Method



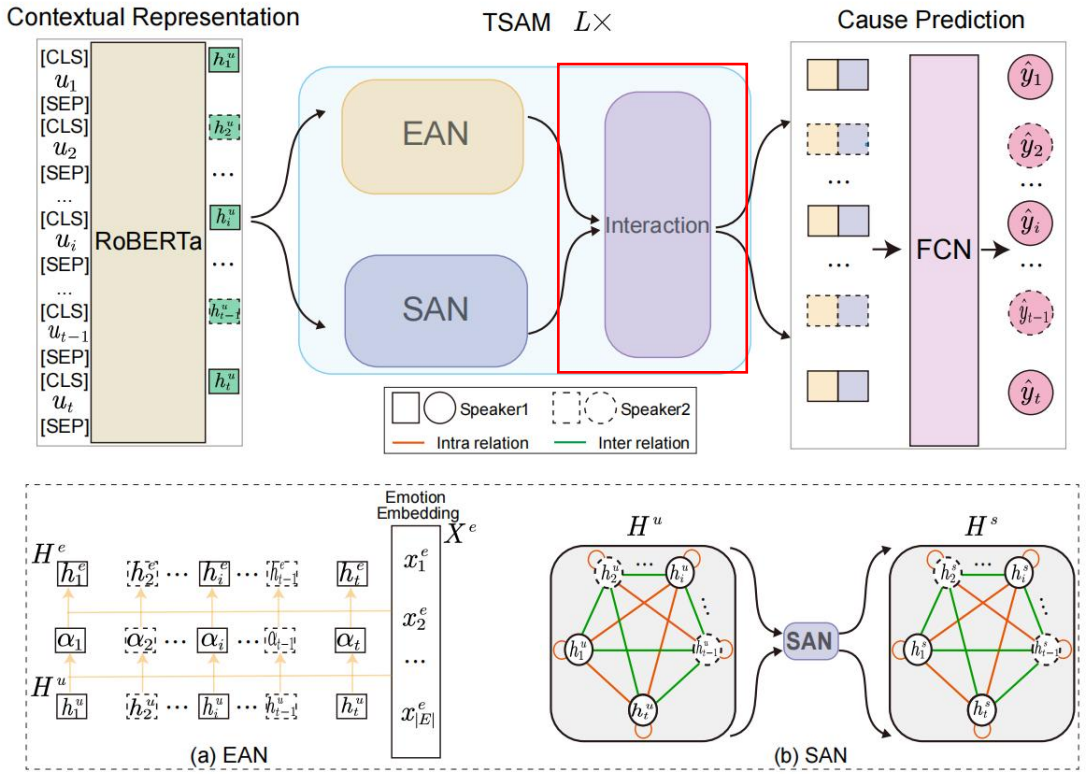
(b) SAN

$$\alpha_{ijr} = softmax_i(LRL(\boldsymbol{a}_r^T[\boldsymbol{W}_r\boldsymbol{h}_i^u; \boldsymbol{W}_r\boldsymbol{h}_j^u])) \quad (7)$$

$$\boldsymbol{h}_{ir} = \sum_{j \in \mathcal{N}^r(i)} \alpha_{ijr}\boldsymbol{W}_r\boldsymbol{h}_j^u \quad (8)$$

$$\boldsymbol{h}_i^s = \sum_{r \in \mathcal{R}} \boldsymbol{h}_{ir} \quad (9)$$

Method



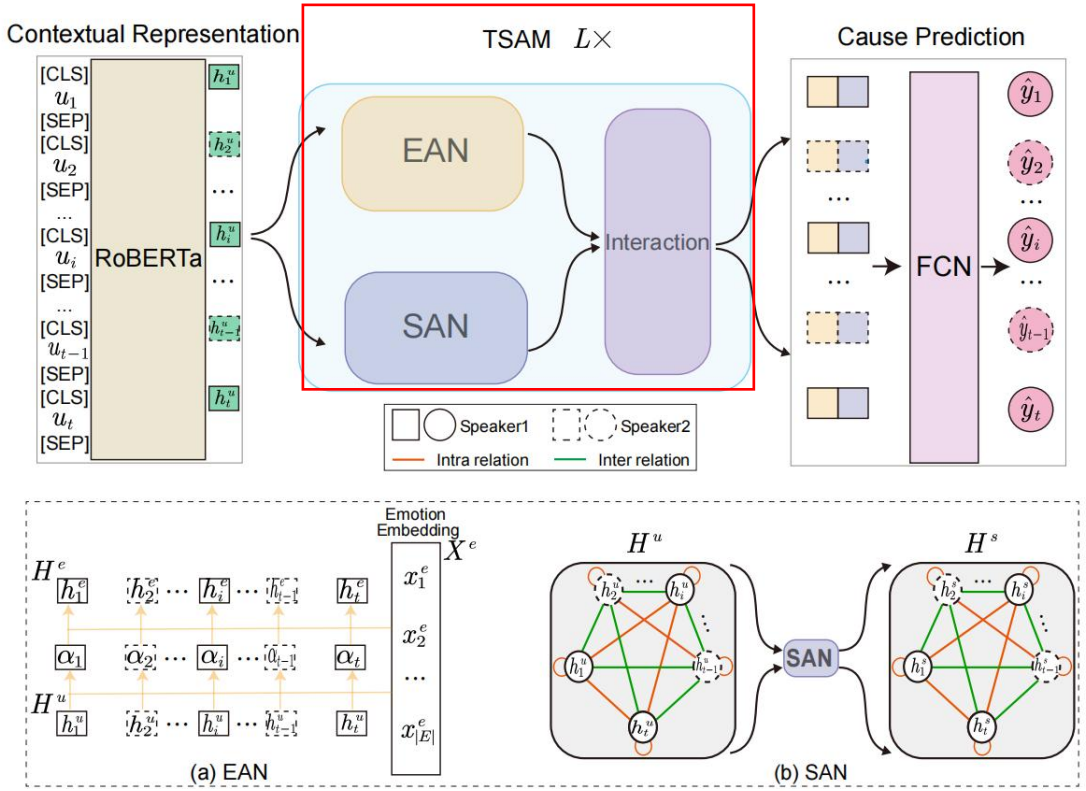
$$\boldsymbol{A}_1 = softmax(\boldsymbol{H}^e \boldsymbol{W}_1 (\boldsymbol{H}^s)^T) \quad (10)$$

$$\boldsymbol{A}_2 = softmax(\boldsymbol{H}^s \boldsymbol{W}_2 (\boldsymbol{H}^e)^T) \quad (11)$$

$$\boldsymbol{H}^{e'} = \boldsymbol{A}_1 \boldsymbol{H}^s \quad (12)$$

$$\boldsymbol{H}^{s'} = \boldsymbol{A}_2 \boldsymbol{H}^e \quad (13)$$

Method



$$\boldsymbol{H}_l^e = \textbf{EAN}(\boldsymbol{E}_l, \boldsymbol{X}^e) \tag{14}$$

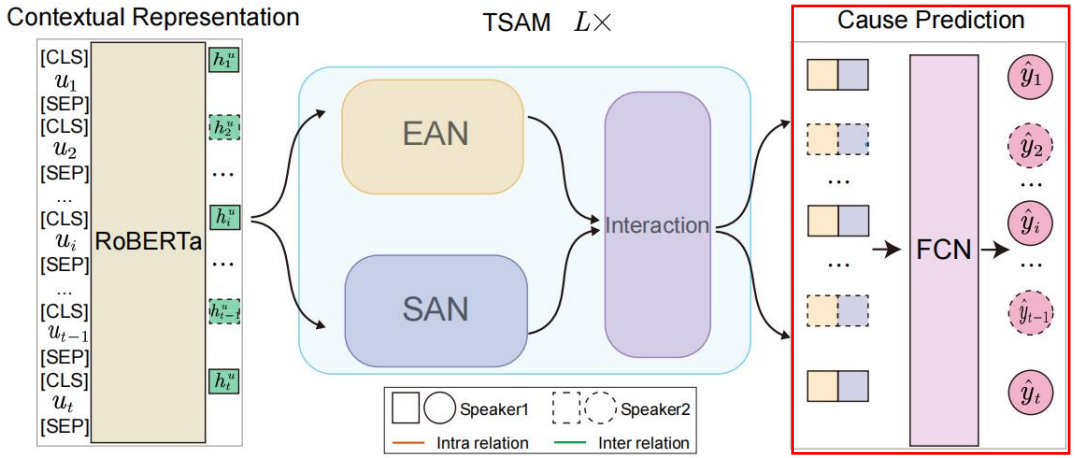
$$\boldsymbol{H}_l^s = \textbf{SAN}(\boldsymbol{S}_l) \tag{15}$$

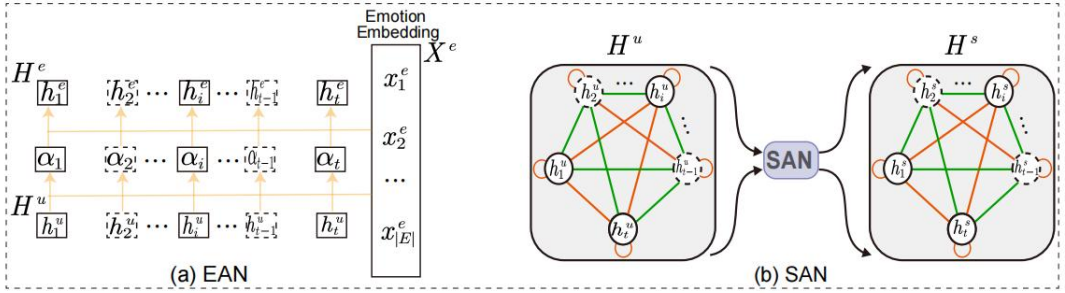
$$\boldsymbol{H}_l^{e'}, \boldsymbol{H}_l^{s'} = \textbf{Interaction}(\boldsymbol{H}_l^e, \boldsymbol{H}_l^s) \tag{16}$$

$$\boldsymbol{E}_{l+1}, \boldsymbol{S}_{l+1} = \boldsymbol{H}_l^{e'}, \boldsymbol{H}_l^{s'} \tag{17}$$

where $\boldsymbol{E}_0 = \boldsymbol{S}_0 = \boldsymbol{H}^u$. The TSAM can be stacked in L layers and $l \in [0, L-1]$.

Method





$$\boldsymbol{l}_i = ReLU(\boldsymbol{W}_1[\boldsymbol{e}_i^L; \boldsymbol{s}_i^L] + \boldsymbol{b}_1) \quad (18)$$

$$\hat{y}_i = sigmoid(\boldsymbol{W}_2\boldsymbol{l}_i + b_2) \quad (19)$$

Experiments

Statistics			RECCON-DD
Data Distributions	Train	Positive	7269
		Negative	20646
	Dev	Positive	347
		Negative	838
	Test	Positive	1894
		Negative	5330
Cause Type Distributions	No Context		43%
	Inter		32%
	Intra		9%
	Hybrid		11%
	Unmentioned		5%

Table 1: Statistics of the RECCON-DD dataset. *No Context*: The cause is present within the target utterance itself; *Inter*: Inter-speaker emotional influences; *Intra*: Intra-speaker emotion influences (Self-Contagion); *Hybrid*: Inter and Intra can jointly cause the emotion of an utterance; *Unmentioned*: Some instances have no explicit emotion causes in the conversational history.

Experiments

#	Model	W/O CH			W/ CH		
		Pos. F1	Neg. F1	macro F1	Pos.F1	Neg.F1	macro F1
0	INDEP_{base}	56.64	85.13	70.88	64.28	88.74	76.51
1	INDEP_{large}	50.48	87.35	68.91	66.23	87.89	77.06
2	JOINT_{base}	-	-	-	66.61	89.11	77.86
3	JOINT_{large}	-	-	-	68.30	89.16	78.73
4	Ours_{base}	-	-	-	68.59	89.75	79.17
5	Ours_{large}	-	-	-	**70.00**†	**90.48**†	**80.24**†

Table 2: Performance of our model and baselines on the test set of RECCON-DD. Bold font denotes the best performance. "Ours" denotes the proposed model without removing any module ("Ours" = "JOINT" + TSAM). "†" denotes that Ours_{large} is statistically significant (Koehn, 2004) better than INDEP_{large} W/ CH (p-value < 0.05).

Experiments

Emotion Information	Pos. F1	Neg. F1	macro F1
No	68.40	89.80	79.10
DAEE	68.90	90.03	79.47
EAN	**70.00**	**90.48**	**80.24**

Table 3: Comparison of different ways of incorporating emotion information. *No*: no emotion information incorporated; *DAEE*: incorporating the emotion information with direct application emotional embedding.

Experiments

Speaker Information	Pos. F1	Neg. F1	macro F1
Not Consider	67.99	89.42	78.71
Consider	**70.00**	**90.48**	**80.24**

Table 4: Results on experiments whether considering speaker information or not in SAN.

Experiments

	Pos. F1	Neg. F1	macro F1
W/O Interaction	68.18	88.93	78.56
W/ Interaction	**70.00**	**90.48**	**80.24**

Table 5: Results on experiments whether removing interaction module or not in TSAM.

Experiments

Models	Intra	Inter
W/O TSAM	62.06	72.67
W/ TSAM	**63.82**	**74.81**

Table 6: Accuracy on the collected samples. *Intra*: Intra-speaker emotional influences; *Inter*: Inter-speaker emotional influences.

Experiments

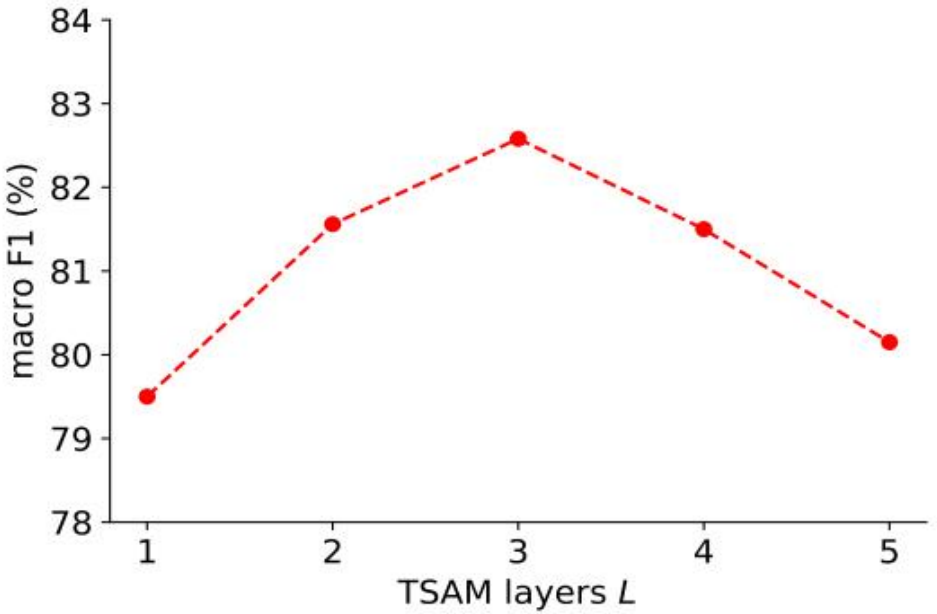


Figure 3: Results of Ours$_{large}$ with various TSAM layers on the development set of RECCON-DD.

Thanks!