Chongqing University of Technology

ATAI
Advanced Technique of Artificial Intelligence

Joint Alignment of Multi-Task Feature and Label Spaces for Emotion Cause Pair Extraction

Shunjie Chen[1], Xiaochuan Shi[1], Jingye Li[1], Shengqiong Wu[2], Hao Fei[2], Fei Li[1†], Donghong Ji[1]

[1]Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, Wuhan, China

[2]School of Computing, National University of Singapore, Singapore

{shunjiechen, shixiaochuan, theodorelee, lifei_csnlp, dhji}@whu.edu.cn
swu@u.nus.edu, haofei37@nus.edu.sg

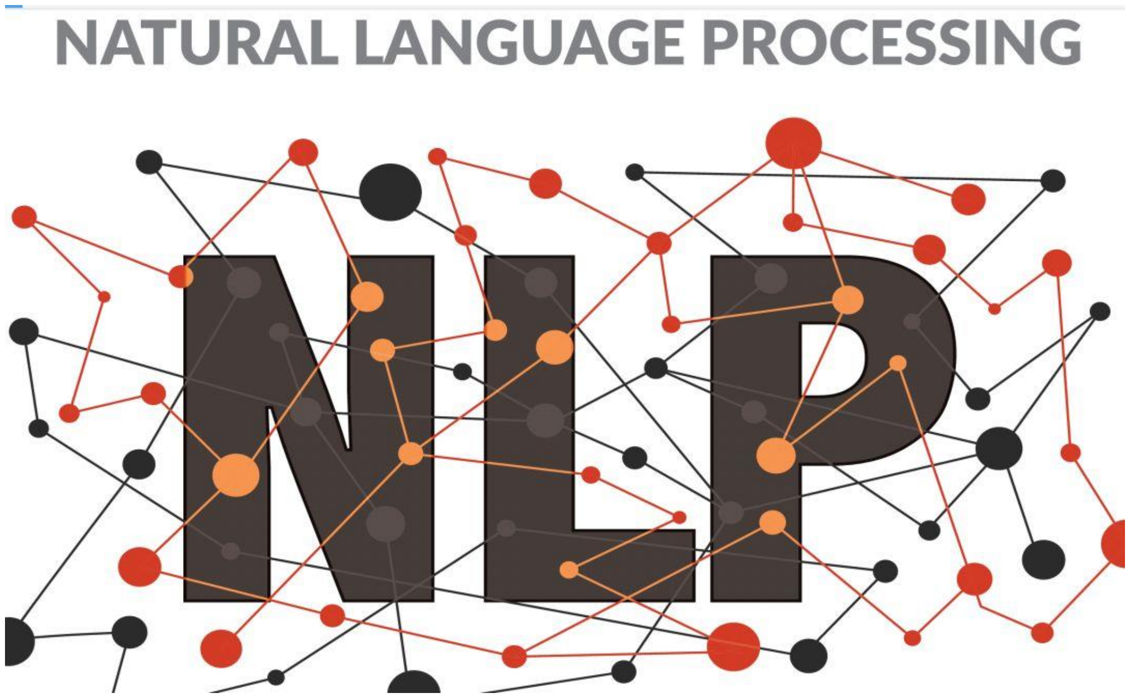
2022.09.28 • ChongQing

2022_COLING

Reported by Yidan Liu

gesis Leibniz-Institut für Sozialwissenschaften

Forschungszentrum · Research Center L3S



1.Introduction

2.Method

3.Experiments

Leibniz-Institut für Sozialwissenschaften

Introduction

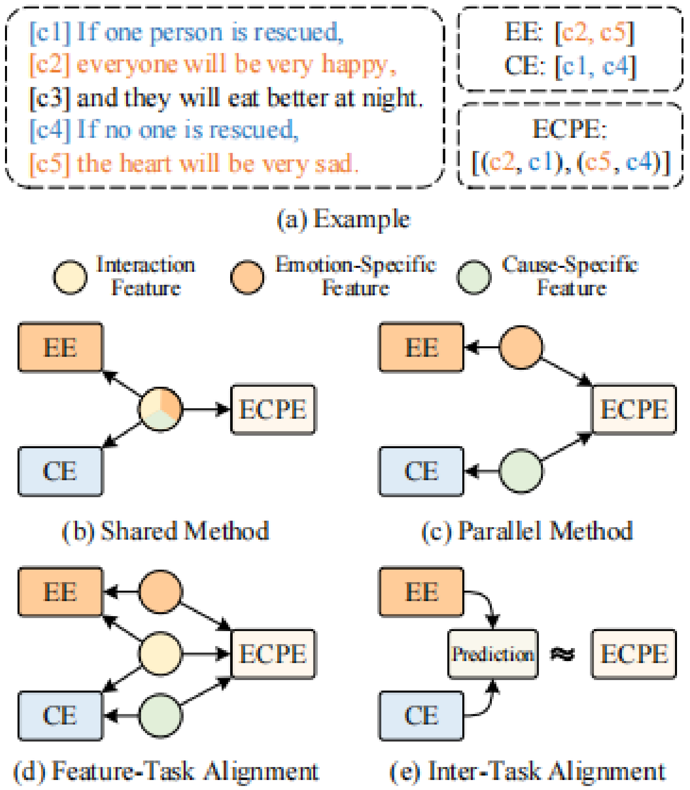


Figure 1: (a) illustrates three subtasks of ECA. (b) and (c) depicts the shared and parallel features encoding method, respectively. In (d) and (e) we show our proposed feature-task alignment mechanism and inter-task alignment mechanism, respectively.

However, existing MTL-based methods either fail to simultaneously model the specific features and the interactive feature in between, or suffer from the inconsistency of label prediction.

Method

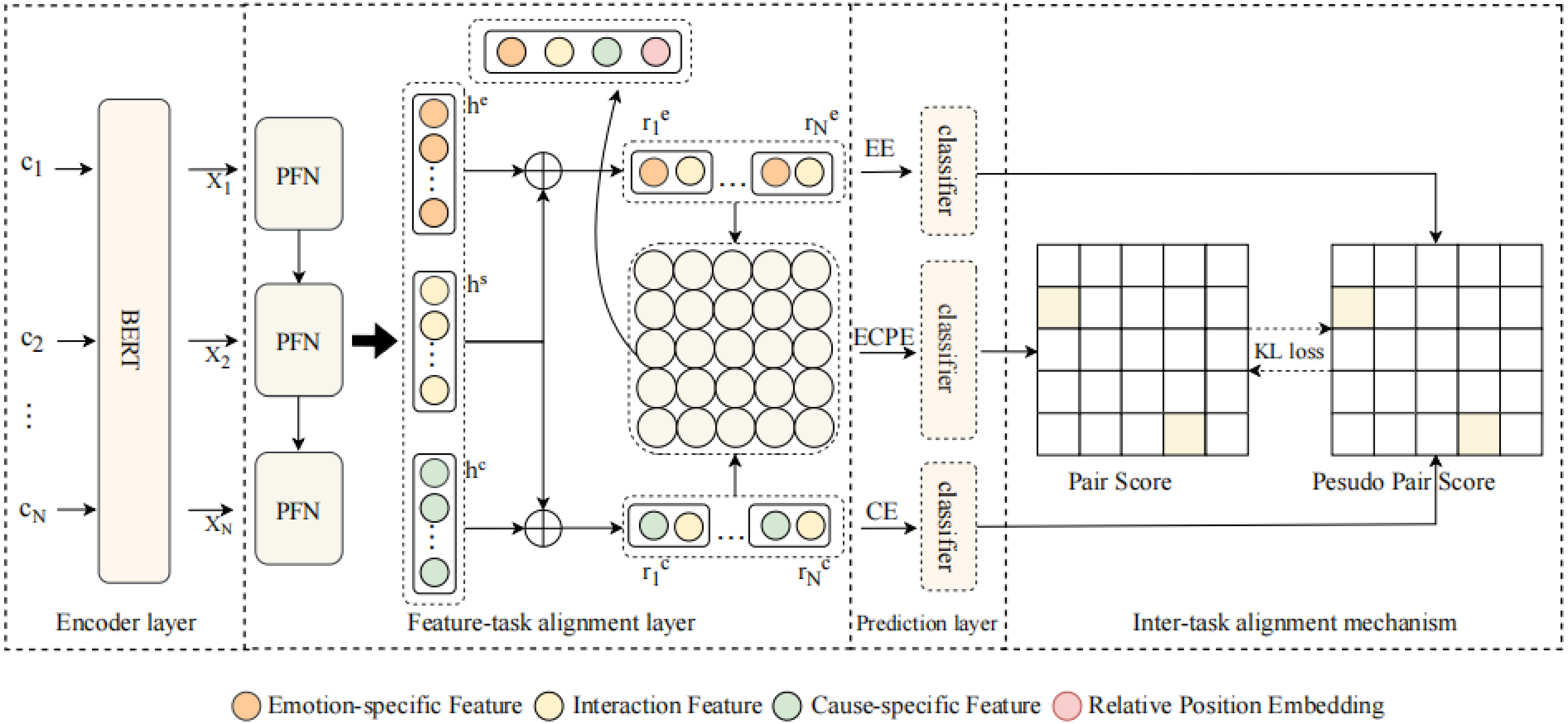
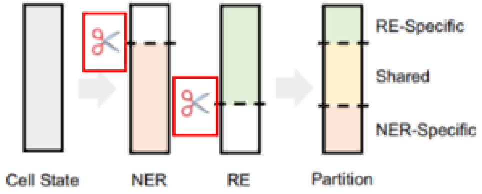


Figure 2: Overview of our A^{2}Net model.

Method

$\tilde{e}$ = cummax(Linear($[x_t; h_{t-1}]$))
$\tilde{r}$ = 1 – cummax(Linear($[x_t; h_{t-1}]$))



Cummax = Softmax + Cumulation

identify cut-off point
e.g. (0.1, 0.1, 0.6, 0.1, 0.1)

segment neurons before and after the cut-off point
e.g. (0.1, 0.2, 0.8, 0.9, 1.0) ≈ (0, 0, 1, 1, 1)

Suppose Entity Gate $\tilde{e}$ = (0, 1, 1) Relation Gate $\tilde{r}$ = (1, 1, 0)

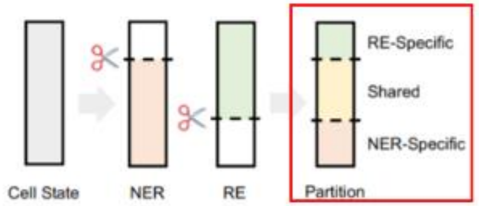
$\rho_{s,c_{t-1}} = \tilde{e}_{c_{t-1}} \circ \tilde{r}_{c_{t-1}}$ → (0, 1, 1) ∘ (1, 1, 0) = (0, 1, 0)

$\rho_{e,c_{t-1}} = \tilde{e}_{c_{t-1}} - \rho_{s,c_{t-1}}$ → (0, 1, 1) – (0, 1, 0) = (0, 0, 1)

$\rho_{r,c_{t-1}} = \tilde{r}_{c_{t-1}} - \rho_{s,c_{t-1}}$ → (1, 1, 0) – (0, 1, 0) = (1, 0, 0)

Gather information from cells

$\rho_e = \rho_{e,c_{t-1}} \circ c_{t-1} + \rho_{e,\tilde{c}_t} \circ c_t$
$\rho_r = \rho_{r,c_{t-1}} \circ c_{t-1} + \rho_{r,\tilde{c}_t} \circ \tilde{c}_t$
$\rho_s = \rho_{s,c_{t-1}} \circ c_{t-1} + \rho_{s,\tilde{c}_t} \circ \tilde{c}_t$



$\mathcal{D} = \{c_1, c_2, \ldots, c_N\}$ $\quad (c_i^e, c_j^c)$

Encoder Layer

$$c_i = \{[CLS], w_{i,1}, w_{i,2}, \ldots, w_{i,M}, [SEP]\}$$

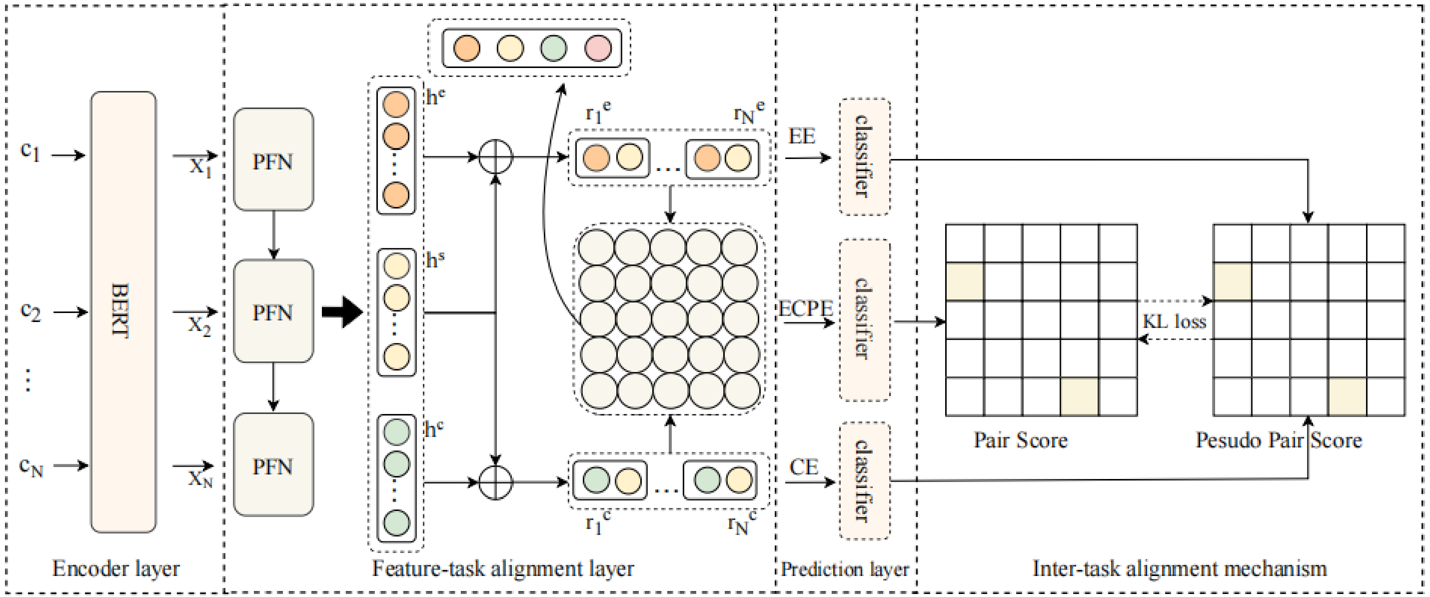
$$\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}.$$

Feature-task Alignment Layer

$$\begin{aligned} \boldsymbol{g}_i^e &= \mathrm{Cummax}(\mathrm{Linear}([\boldsymbol{x}_i; \boldsymbol{h}_{i-1}])), \\ \boldsymbol{g}_i^c &= 1 - \mathrm{Cummax}(\mathrm{Linear}([\boldsymbol{x}_i; \boldsymbol{h}_{i-1}])), \end{aligned} \quad (1)$$

$$\begin{aligned} \boldsymbol{f}_i^s &= \boldsymbol{g}_i^e \circ \boldsymbol{g}_i^c, \\ \boldsymbol{f}_i^e &= \boldsymbol{g}_i^e - \boldsymbol{f}_i^s, \\ \boldsymbol{f}_i^c &= \boldsymbol{g}_i^c - \boldsymbol{f}_i^s, \end{aligned} \quad (2)$$

Method



Emotion-specific Feature   Interaction Feature   Cause-specific Feature   Relative Position Embedding

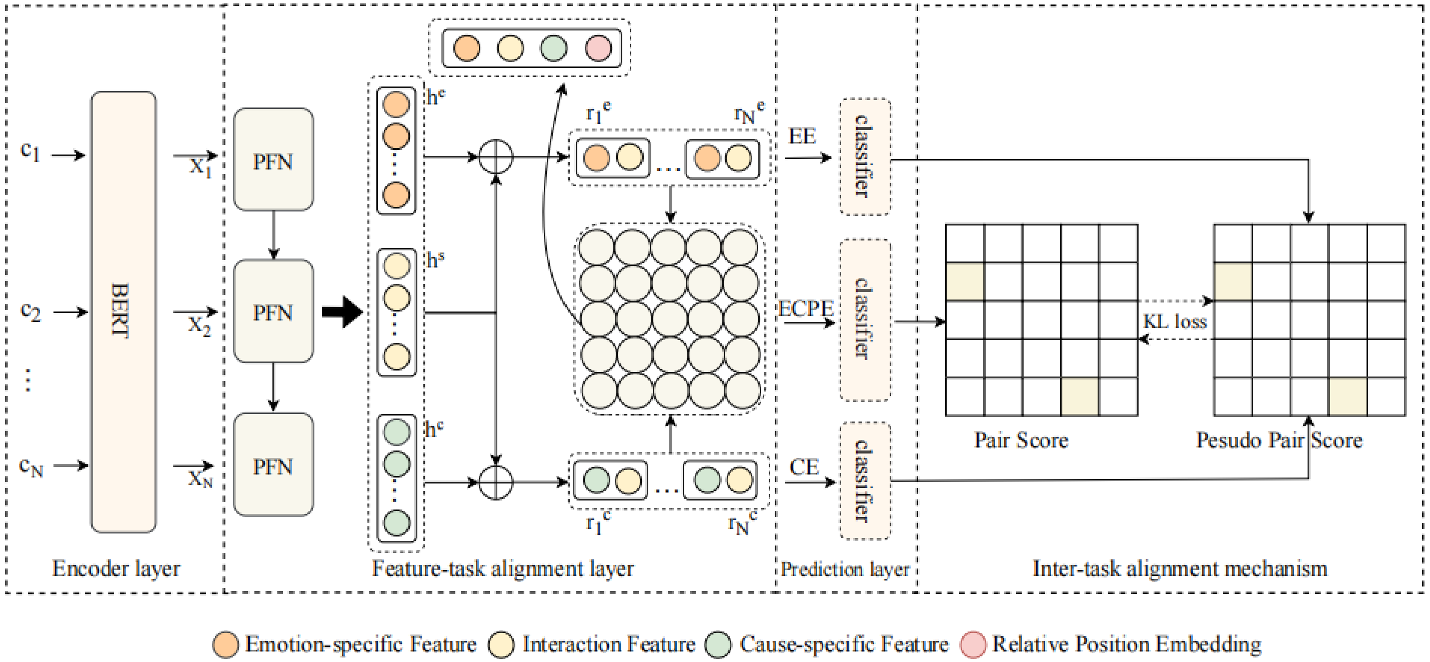
$$\begin{aligned} \widetilde{c}_i &= \tanh(\mathrm{Linear}([x_i; h_{i-1}])), \\ p_i^s &= f_i^s \circ c_{i-1} + o_i^s \circ \widetilde{c}_i, \\ p_i^e &= f_i^e \circ c_{i-1} + o_i^e \circ \widetilde{c}_i, \\ p_i^c &= f_i^c \circ c_{i-1} + o_i^c \circ \widetilde{c}_i, \end{aligned} \tag{3}$$

$$\begin{aligned} h_i^s &= \tanh(p_i^s), \\ h_i^e &= \tanh(p_i^e), \\ h_i^c &= \tanh(p_i^c). \end{aligned} \tag{4}$$

$$\begin{aligned} c_i &= \mathrm{Linear}([p_i^e; p_i^s; p_i^c]), \\ h_i &= \tanh(c_i). \end{aligned} \tag{5}$$

$$\begin{aligned} h_i^{e'} &= h_i^e + h_i^s, \\ h_j^{c'} &= h_j^c + h_j^s, \\ r_{ij} &= [h_i^{e'}; h_j^{c'}; e_{ij}], \end{aligned} \tag{6}$$

Method



$$\boldsymbol{r}_i^e = [\boldsymbol{h}_i^e; \boldsymbol{h}_i^s] \qquad \boldsymbol{r}_i^c = [\boldsymbol{h}_i^c; \boldsymbol{h}_i^s]$$

Prediction Layer

$$\begin{aligned} \hat{y}_i^e &= \sigma(\mathrm{FFN}(\boldsymbol{r}_i^e)) \,, \\ \hat{y}_i^c &= \sigma(\mathrm{FFN}(\boldsymbol{r}_i^c)) \,, \end{aligned} \tag{7}$$

$$\mathcal{L}_{aux} = -\sum_{i=1}^{N} (y_i^e \log(\hat{y}_i^e) + y_i^c \log(\hat{y}_i^c)) \,, \tag{8}$$

$$\hat{y}_{ij}^p = \sigma(\mathrm{FFN}(\boldsymbol{r}_{ij})) \,. \tag{9}$$

$$\mathcal{L}_{pair} = -\sum_{i=1}^{N} \sum_{j=1}^{N} y_{ij}^p \log(\hat{y}_{ij}^p) \,, \tag{10}$$

Method

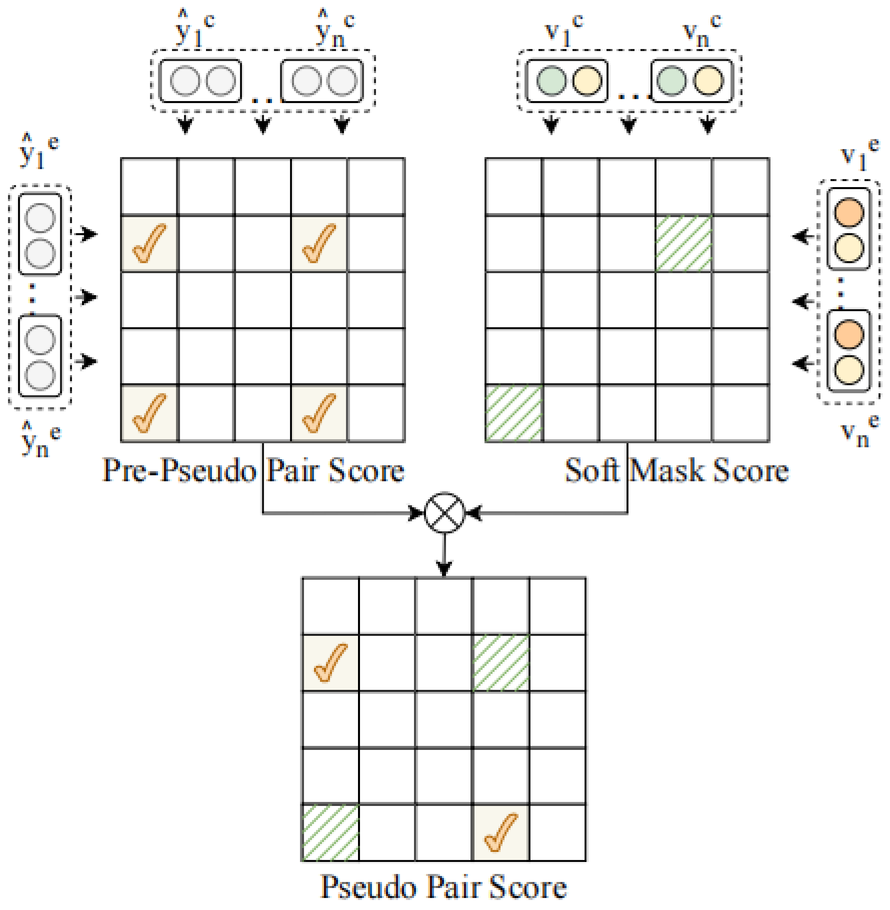


Figure 3: The generation of pseudo pair score, where ✓ denotes candidate emotion-cause pairs, green grids represent the masked pair.

Inter-task Alignment Mechanism

$$\tilde{y}_{ij}^{p} = \alpha_{ij}\sqrt{\hat{y}_i^e \hat{y}_j^c}\,, \tag{11}$$

$$t_{ij} = \frac{(v_i^e)^\top v_j^c}{\sqrt{d}}\,, \quad \alpha_{ij} = \frac{\exp(t_{ij})}{\sum_j^N \exp(t_{ij})}\,, \tag{12}$$

where v_i^e and v_j^c are obtained from $r_i^e = [h_i^e; h_i^s]$ and $r_j^c = [h_j^c; h_j^s]$ with FFNs, respectively. The d denotes the dimension of v_i^e and v_j^c.

$$\mathcal{L}_{KL} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\mathrm{KL}(\tilde{y}_{ij}^p||\hat{y}_{ij}^p) + \mathrm{KL}(\hat{y}_{ij}^p||\tilde{y}_{ij}^p)) = \frac{1}{2}(\sum_{i=1}^{N}\sum_{j=1}^{N}(\tilde{y}_{ij}^p \log(\frac{\tilde{y}_{ij}^p}{\hat{y}_{ij}^p}) + \hat{y}_{ij}^p \log(\frac{\hat{y}_{ij}^p}{\tilde{y}_{ij}^p}))\,. \tag{13}$$

$$\mathcal{L} = \mathcal{L}_{pair} + \lambda_1\mathcal{L}_{aux} + \lambda_2\mathcal{L}_{KL}\,, \tag{14}$$

Experiment

Approach	ECPE			EE			CE		
	P	R	F1	P	R	F1	P	R	F1
ANTS	72.43	63.66	67.76	81.96	73.29	77.39	74.90	66.02	70.18
TransECPE	73.74	63.07	67.99	87.16	82.44	84.74	75.62	64.71	69.74
ECPE-2D	72.92	65.44	68.89	86.27	**92.21**	89.10	73.36	69.34	71.23
PairGCN	76.92	67.91	72.02	88.57	79.58	83.75	<u>79.07</u>	69.28	73.75
RANKCP	71.19	<u>76.30</u>	73.60	<u>91.23</u>	89.99	<u>90.57</u>	74.61	77.88	76.15
ECPE-MLL	<u>77.00</u>	72.35	74.52	86.08	<u>91.91</u>	88.86	73.82	<u>79.12</u>	76.30
MGSAG	**77.43**	73.21	<u>75.21</u>	**92.08**	82.11	87.17	**79.79**	74.68	<u>77.12</u>
A^2Net(ours)	75.03	**77.80**	**76.34**	90.67	90.98	**90.80**	77.62	**79.20**	**78.35**

Table 1: Comparisons with baselines on Chinese benchmark ECPE corpus. For a fair comparison, they all use BERT as the encoder. The best performance is in **bold** and the second best performance is <u>underlined</u>.

Experiment

	ECPE	EE	CE
A^2Net (ours)	**76.34**	**90.80**	**78.35**
w/ Shared encoding	69.97	84.81	72.66
w/ Parallel encoding	75.59	89.75	78.03

Table 2: Performances (F1) with different feature encoding schemes.

	ECPE	EE	CE
A^2Net (ours)	**76.34**	**90.80**	**78.35**
w/o ECPE→EE×CE	75.83	90.65	78.05
w/o EE×CE→ECPE	75.50	90.54	77.60
w/o ITA	75.32	90.05	77.37
w/o EE & CE	74.39	-	-

Table 3: Ablation study of inter-task alignment module and auxiliary task (F1). The ECPE→EE×CE means we use the prediction distribution of ECPE to align to EE×CE (i.e., $\mathrm{KL}(\tilde{y}^p_{ij}||\hat{y}^p_{ij})$ in Eq.13), and vice versa.

Experiment

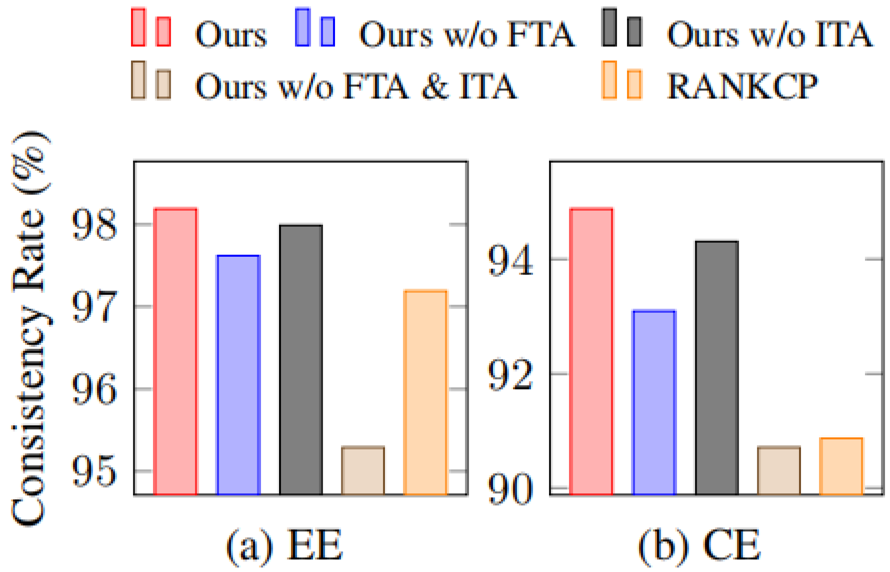


Figure 4: Consistency of ECPE and EE (a), as well as CE (b).

Model	#Param	Speed(doc/s)
RANKCP	105.97M	195
A^2Net (ours)	104.97M	195
w/o ITA	104.97M	195

Table 5: Parameter number and inference speed comparisons on ECPE. All models are tested with batch size 4.

Experiment

Document	A^{2}Net(w/o ITA)	A^{2}Net	Ground-truth
... The police visited the villagers of Nanyuan Village (c3), and they learned that Meng was playing mahjong at a mahjong parlor opposite his home the day before the incident (c4), through inquiries (c5), it wa- -s found that only Wang from the same village had gone out to an unk- -nown destination(c6), which aroused the suspicion of the police (c7).	ECPE:[c7, c6] EE:[] CE:[c6]	ECPE:[c7, c6] EE:[c7] CE:[c6]	ECPE:[c7, c6] EE:[c7] CE:[c6]
On March 14 (c1), a magnitude 4.3 earthquake occurred in Yingquan District, Fuyang City, Anhui (c2). Then (c3), a rumor of a magnitude 6.8 earthquake occurred in Fuyang City at 2:15 am on March 15. (c4), which caused people to panic (c5)...	ECPE:[c5, c4] EE:[c5] CE:[c2], [c4]	ECPE:[c5, c4] EE:[c5] CE:[c4]	ECPE:[c5, c4] EE:[c5] CE:[c4]
Mr. Feng said frankly (c1), Jingjing is naughty on weekdays (c2), and sometimes he is not polite (c3), but when it comes to the reason for th- -is injury(c4), he can't hide his anger (c5), just because of my son Dra- -nk other children's yogurt (c6). Teacher Xing lost her mind (c7), she was emotionally out of control (c8), then pulled the child out of the do- -or (c9), the child was injured when the door was closed (c10)...	ECPE:[c5,c4],[c5,c6] EE:[c5] CE:[c4], [c6]	ECPE:[c5,c4] EE:[c5] CE:[c4]	ECPE:[c5,c4] EE:[c5] CE:[c4]

Table 4: Two examples for the case study. The words in orange are the emotion clause, and the words in blue are the cause clause. The green means correct predictions, red means wrong predictions.

Thank you!