Artificial

Joint Constrained Learning with Boundary-adjusting for Emotion-Cause Pair Extraction

Huawen Feng, Junlong Liu, Junhao Zheng, Haibin Chen, Xichen Shang, Qianli Ma*
School of Computer Science and Engineering,
South China University of Technology, Guangzhou, China
541119578@qq.com, qianlima@scut.edu.cn

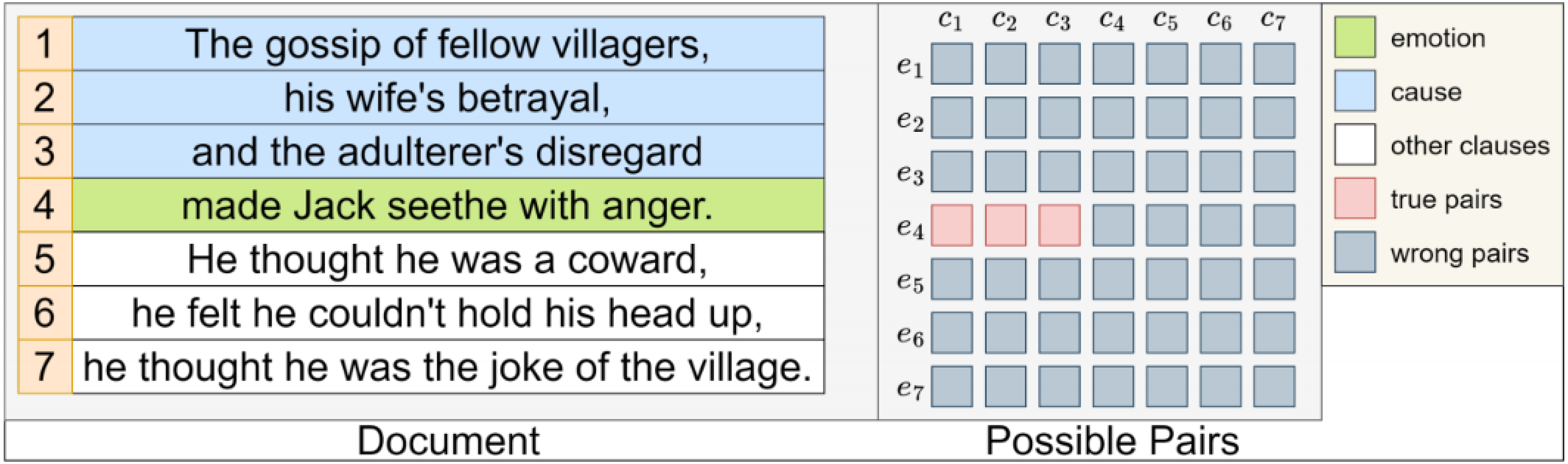
2023.9.14 • ChongQing

—— ACL 2023

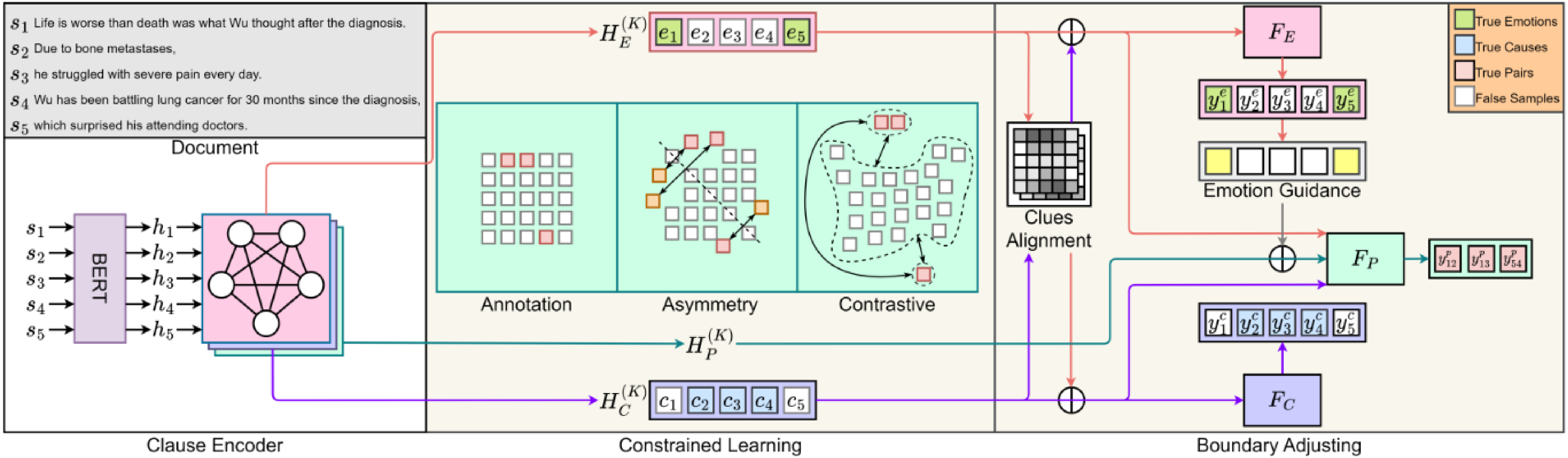
Leibniz-Institut für Sozialwissenschaften

Reported by Renhui Luo

Introduction



Overview

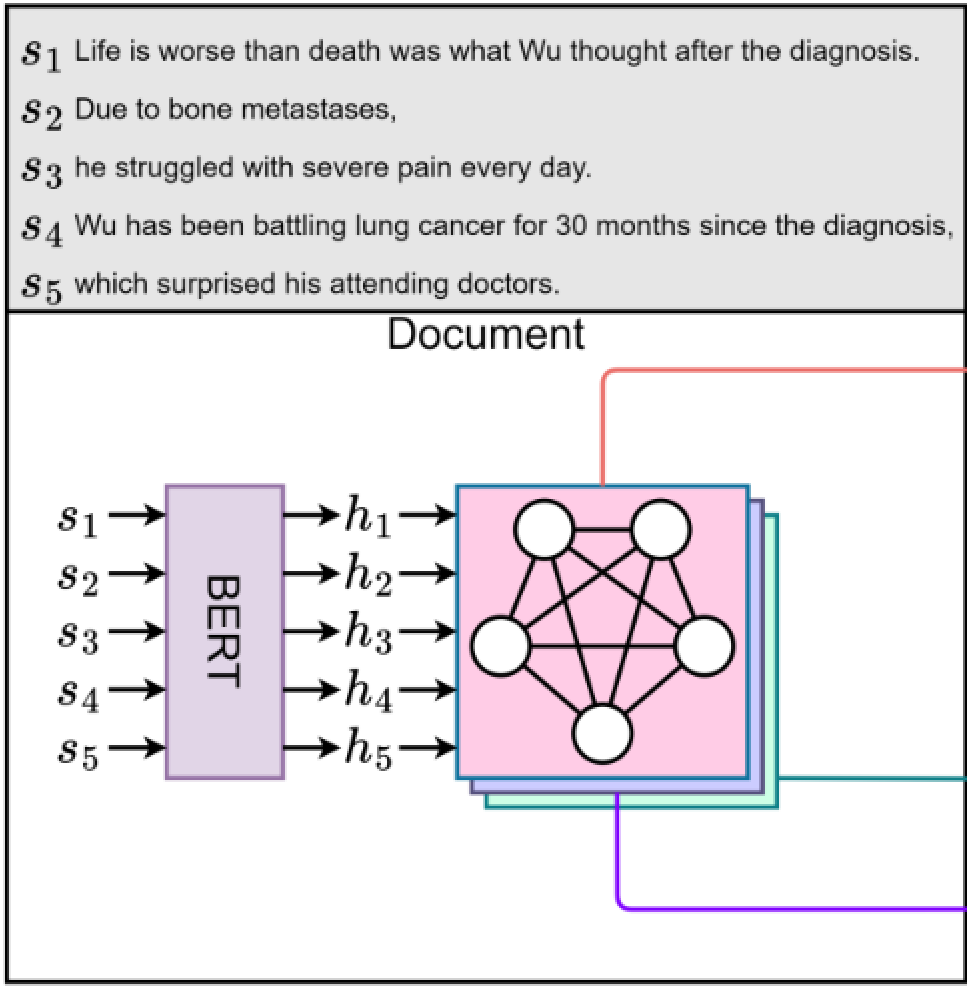


$$P = \{..., (s_i, s_j), ...\} \qquad i, j \in [1, n] \qquad (1)$$

$$Y_i^e = \begin{cases} 1 & if(s_i, s_j) \in P \\ 0 & otherwise \end{cases} \qquad (2)$$

$$Y_j^c = \begin{cases} 1 & if(s_i, s_j) \in P \\ 0 & otherwise \end{cases} \qquad (3)$$

Method

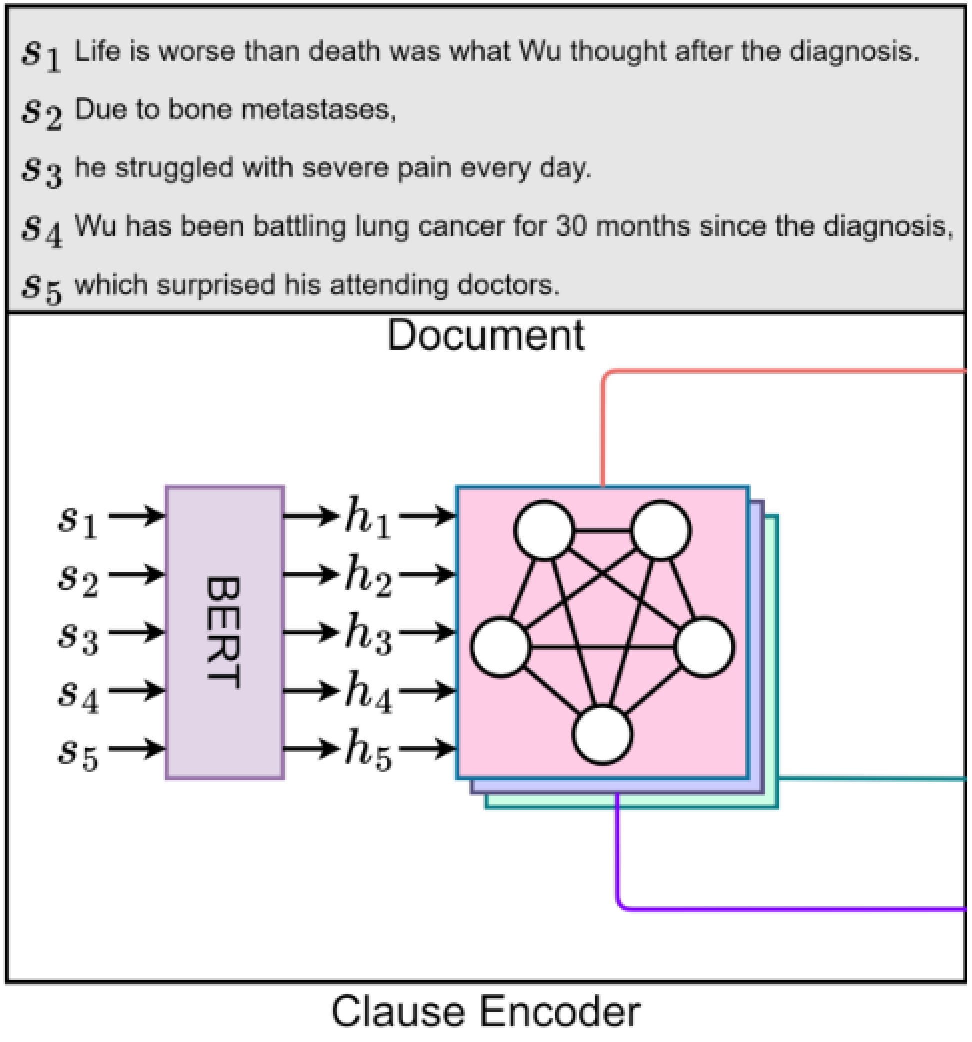


Clause Encoder

$$
\begin{aligned}
H_E^{(0)} &= [h_1^{e(0)}, h_2^{e(0)}, \ldots, h_n^{e(0)}] \\
H_C^{(0)} &= [h_1^{c(0)}, h_2^{c(0)}, \ldots, h_n^{c(0)}] \\
H_P^{(0)} &= [h_{11}^{p(0)}, h_{12}^{p(0)}, \ldots, h_{nn}^{p(0)}] \\
h_i^{e(0)} &= h_i^{c(0)} = h_i \\
h_{ij}^{p(0)} &= Linear_{pair}([h_i; h_j])
\end{aligned}
\tag{4}
$$

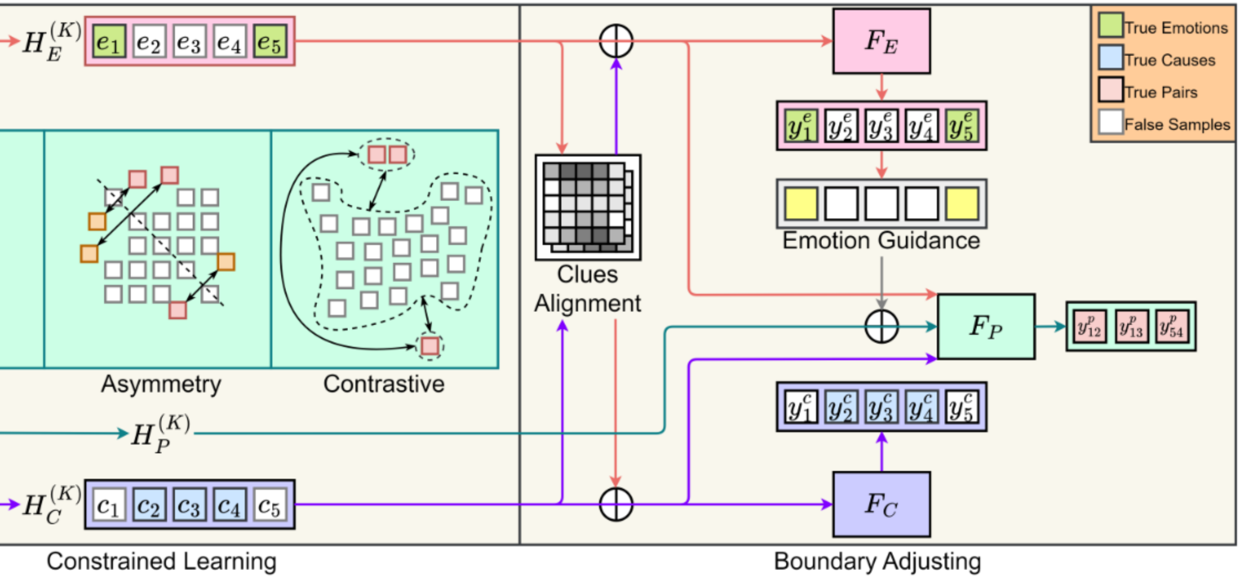
$$
h_v^{(t+1)} = (W^{(t)} h_v^{(t)} + b^{(t)}) + \frac{1}{|N(v)|} \sum_{r \in R} \sum_{z \in N(v)} (W_r^{(t)} h_z^{(t)} + b_r^{(t)}) \tag{5}
$$

Method



$$
\begin{aligned}
H_E^{(K)} &= [e_1, e_2, \ldots, e_n] \\
H_C^{(K)} &= [c_1, c_2, \ldots, c_n] \\
H_P^{(K)} &= [p_{11}, p_{12}, \ldots, p_{nn}] \\
e_i = h_I^{e(K)} \quad & c_i = h_I^{c(K)} \quad p_{ij} = h_{ij}^{p(K)}
\end{aligned}
\tag{6}
$$

Method

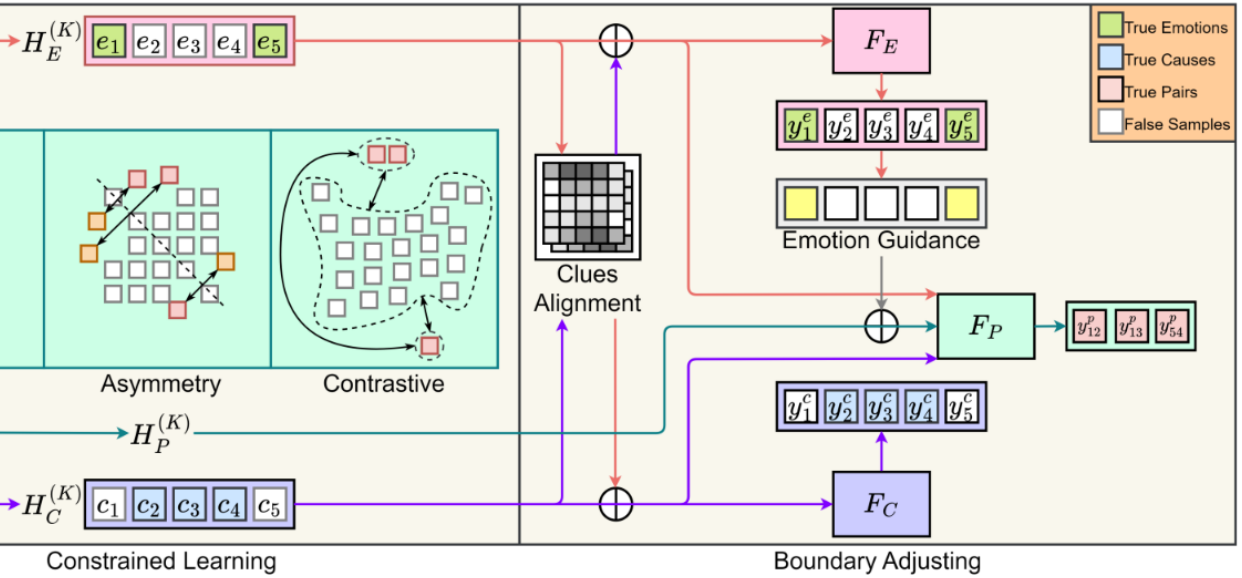


$$
\begin{gathered}
m_{ij} = (c_i)^T \times e_j \\
c_i \in H_C^{(K)} \quad e_j \in H_E^{(K)} \\
M_{ij}^{E2C} = \frac{exp(m_{ij})}{\sum_{k=1}^{n} exp(m_{ik})}
\end{gathered} \tag{10}
$$

$$
\begin{gathered}
U^{E2C} = [u_1^{E2C}, u_2^{E2C}, ..., u_n^{E2C}] \\
u_i^{E2C} = \sum_{j=1}^{n} (M_{ij}^{E2C} \cdot e_j)
\end{gathered} \tag{11}
$$

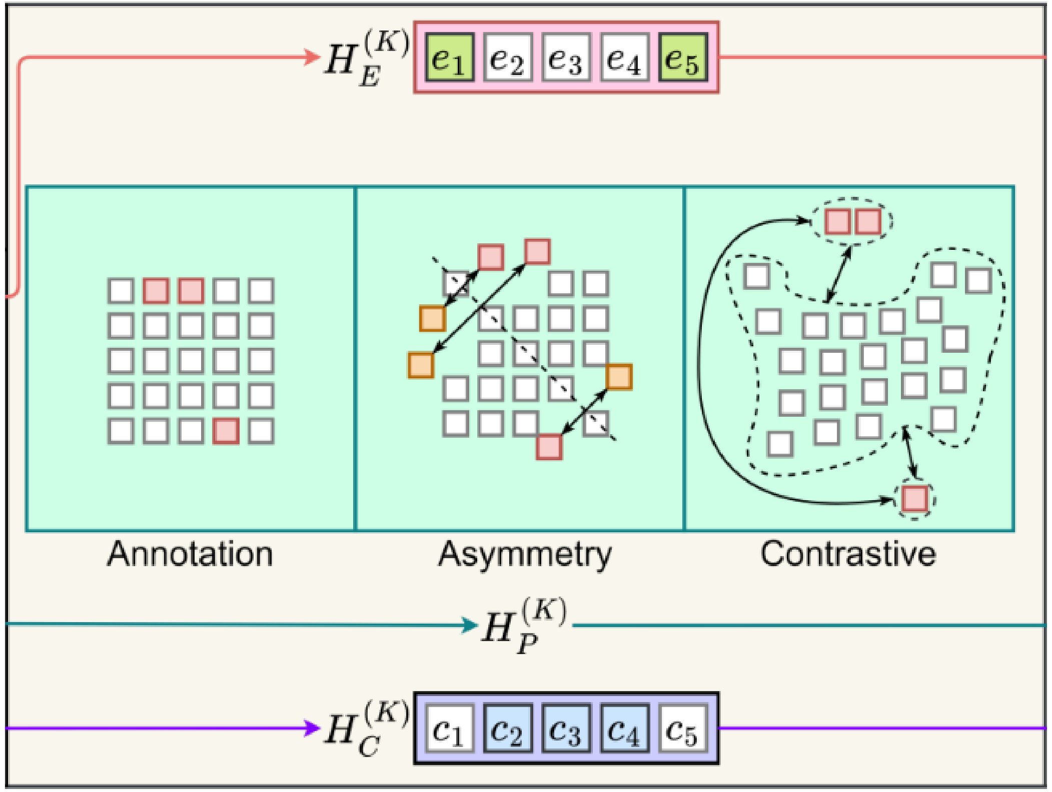
$$
\begin{gathered}
\overline{H_C} = H_C^{(K)} + ReLU(W_{e2c}U^{E2C} + b_{e2c}) \\
Y^C = F_C(\overline{H_C})
\end{gathered} \tag{12}
$$

Method



$$
\begin{aligned}
Y^P &= F_P(\overline{H_P}) \\
\overline{H_P} &= [\overline{p_{11}}, \overline{p_{12}}, ..., \overline{p_{nn}}] \\
\overline{p_{ij}} &= W_p ReLU(p_{ij} + EMB_e(Y_i^e)) + b_p \\
p_{ij} &\in H_P^{(K)}
\end{aligned}
\tag{13}
$$

Method



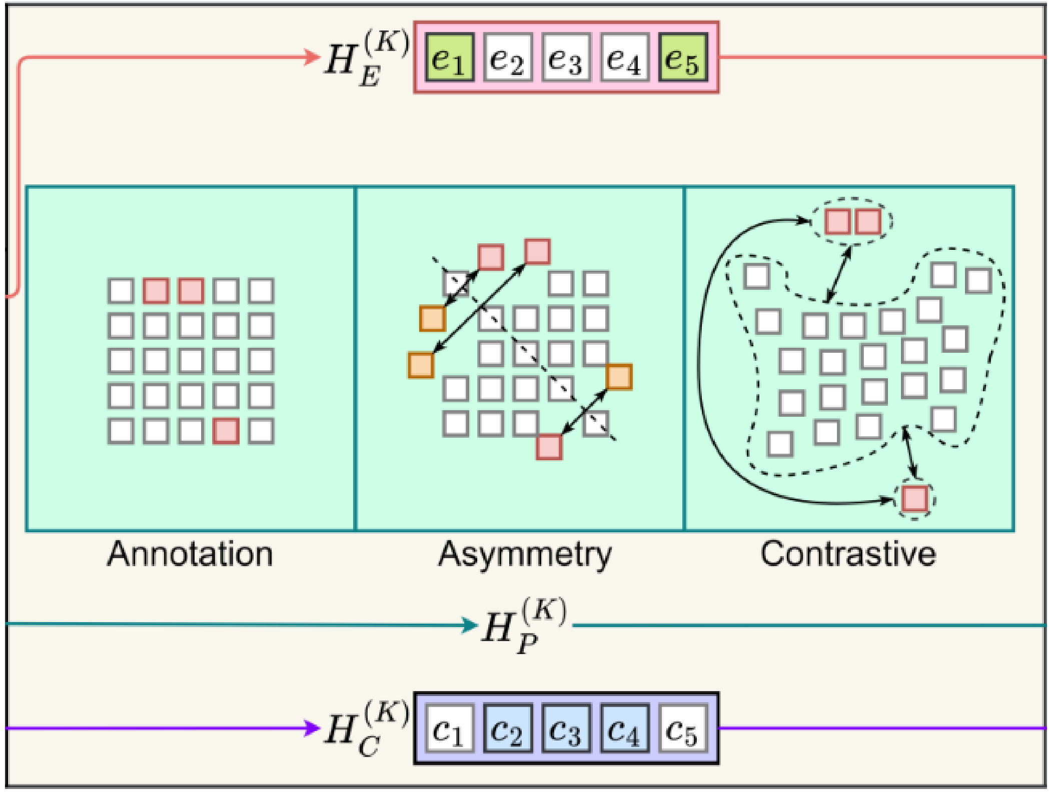
Constrained Learning

$$L_{Annotation} = \sum_{(s_i,s_j)\in\hat{P}} -log(y_{ij}^p) \tag{7}$$

$$L_{Asymmetry} = \sum_{(s_i,s_j)\in\hat{P}} log(y_{ji}^p) - log(y_{ij}^p) \tag{8}$$

$$L_{Contrastive} = \frac{1}{|\hat{P}|} \sum_{(s_i,s_j)\in\hat{P}} max(d(p_{ij}, center_i) - d(p_{ij}, x_{ij}) + \gamma, 0) \tag{9}$$

Method



Constrained Learning

$$
\begin{aligned}
L &= L_{emotion} + L_{cause} + L_{Annotation} \\
&\quad + \alpha L_{Asymmetry} + \beta L_{Contrastive} \\
L_{emotion} &= -\frac{1}{|D|}\sum_{i=1}^{|D|} \hat{Y}_i^e \log y_i^e \\
L_{cause} &= -\frac{1}{|D|}\sum_{i=1}^{|D|} \hat{Y}_i^c \log y_i^c
\end{aligned}
\tag{14}
$$

Experiments

Item	Number	Percentage(%)
documents	1,945	100
-w/ 1 EC pair	1,746	89.8
-w/ 2 EC pairs	177	9.1
-w/ 3 EC pairs	22	1.1
pairs	490,367	100
-EC pairs	2,167	0.4
-non EC pairs	488,200	99.6

Table 1: Detailed dataset statistics.

Config	Value
Device	GeForce RTX 3090
Platform	Pytorch 1.8.0
Backbone	BERT-base-Chinese
Dimension	768
Batch Size	4
Epochs	50
Learning Rate	2e-5
Warmup Proportion	0.1
Dropout	0.2
K	1
α	0.15
β	0.5

Experiments

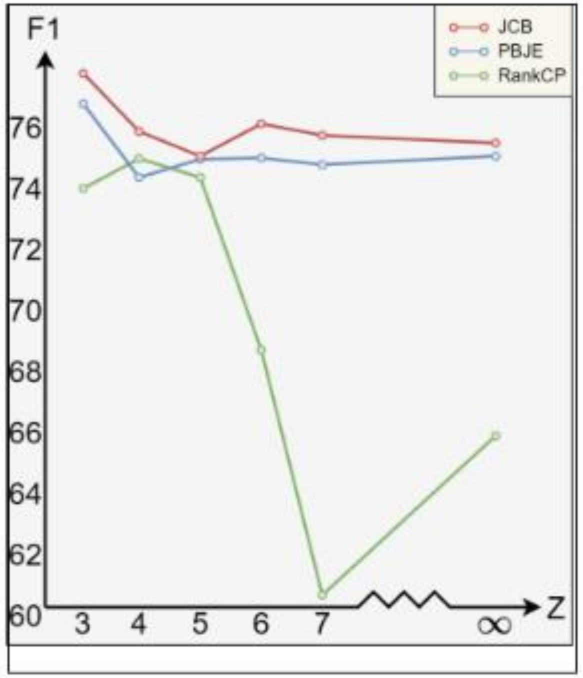


Figure 3: The fluctuation of performance when relative distance changes.

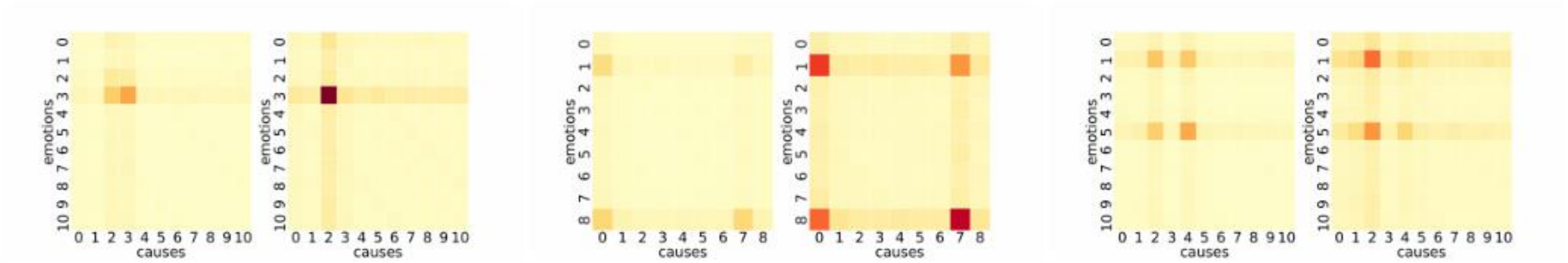
Experiments

Models	Pair Extraction			Emotion Extraction			Cause Extraction		
	P	R	F1	P	R	F1	P	R	F1
ECPE-2D	72.92	65.44	68.89	86.27	92.21[#1]	89.10	73.36	69.34	71.23
TransECPE	77.08	65.32	70.72	88.79	83.15	85.88	78.74	66.89	72.33
PairGCN	76.92	67.91	72.02	88.57	79.58	83.75	79.07	68.28	73.75
UTOS	73.89	70.62	72.03	88.15	83.21	85.56	76.71	73.20	74.71
MTST-ECPE	75.78	70.51	72.91	85.83	80.94	83.21	77.64	72.36	74.77
RankCP	71.19	76.30[#1]	73.60	91.23[#1]	89.99	90.57[#1]	74.61	77.88[#2]	76.15
ECPE-MLL	77.00	72.35	74.52	86.08	91.91[#2]	88.86	73.82	79.12[#1]	76.30
PBJE	79.22[#1]	73.84	76.37[#2]	90.77[#2]	86.91	88.76	81.79[#1]	76.09	78.78[#2]
JCB	79.10[#2]	75.84[#2]	77.37[#1]	90.77[#2]	87.91	89.30[#2]	81.41[#2]	77.47	79.34[#1]

Experiments

Models	Pair Extraction			Emotion Extraction			Cause Extraction		
	P	R	F1	P	R	F1	P	R	F1
JCB	79.10	75.84	77.37	90.77	87.91	89.30	81.41	77.47	79.34
-w/o Asymmetry Constraint	78.82	74.13	76.34	90.91	87.20	88.99	80.71	75.79	78.11
-w/o Contrastive Constraint	76.83	75.42	76.05	88.72	87.54	88.08	80.02	77.23	78.54
-w/o Constrained Learning	76.31	74.37	75.26	90.45	88.71	89.53	79.58	76.34	77.88
-w/o Emotion Clues	78.93	74.38	76.55	91.16	87.77	89.41	81.02	76.18	78.50
-w/o Cause Clues	79.20	74.44	76.67	91.01	87.49	89.16	81.28	76.33	78.66
-w/o Clues Alignment	79.64	73.46	76.38	91.30	86.62	88.87	81.45	75.25	78.19
-w/o Emotion Guidance	78.20	75.50	76.76	90.80	88.29	89.50	80.67	76.98	78.74
-w/o Boundary Adjusting	78.32	74.32	76.19	90.86	87.49	89.10	81.17	76.36	78.61
Clause Encoder (BERT+GCN)	73.01	76.23	74.44	89.17	88.77	88.92	77.25	78.21	77.62

Experiments



(a) Label: (s_3, s_2).

(b) Label: (s_1, s_0) and (s_8, s_7).

(c) Label: (s_1, s_2).

Figure 4: The heat maps of the output of PBJE (left graphs) and JCB (right graphs). The deeper color means the higher confidence. Three subfigures show asymmetric output, differentiated output, and accurate match of JCB compared with PBJE.

Experiments

Models	Pair Extraction		
	P	R	F1
RankCP	64.26(6.93↓)	66.94(9.36↓)	65.49(8.11↓)
PBJE	78.41(0.81↓)	71.31(2.53↓)	74.66(1.71↓)
JCB	78.93(0.17↓)	71.68(4.16↓)	75.09(2.28↓)

Table 5: The results of RankCP, PBJE, and JCB without the relative distance constraint.

Experiments

Models	Pair Extraction		
	P	R	F1
$k = 1$	79.10	75.84	77.37
$k = 2$	78.27	73.16	75.58
$k = 3$	76.99	72.67	74.7

Table 6: The decrease of performance with the increase of k.

Thanks!