HCL-TAT: A Hybrid Contrastive Learning Method for Few-shot Event Detection with Task-Adaptive Threshold

Ruihan Zhang[1,2], **Wei Wei**[1,2]*, **Xian-Ling Mao**[3], **Rui Fang**[4], **Dangyang Chen**[4]
[1]Cognitive Computing and Intelligent Information Processing (CCIIP) Laboratory, School of Computer Science and Technology, Huazhong University of Science and Technology
[2]Joint Laboratory of HUST and Pingan Property & Casualty Research (HPL)
[3]Department of Computer Science and Technology, Beijing Institute of Technology
[4]Ping An Property & Casualty Insurance company of China, Ltd
ruihanzhang@hust.edu.cn, weiw@hust.edu.cn, maoxl@bit.edu.cn
fangrui051@pingan.com.cn, chendangyang273@pingan.com.cn

（EMNLP-2022)

gesis Leibniz-Institut für Sozialwissenschaften

Forschungszentrum · Research Center L3S

Reported by Zhaoze Gao

1. Introduction
2. Approach
3. Experiments

gesis
Leibniz-Institut für Sozialwissenschaften

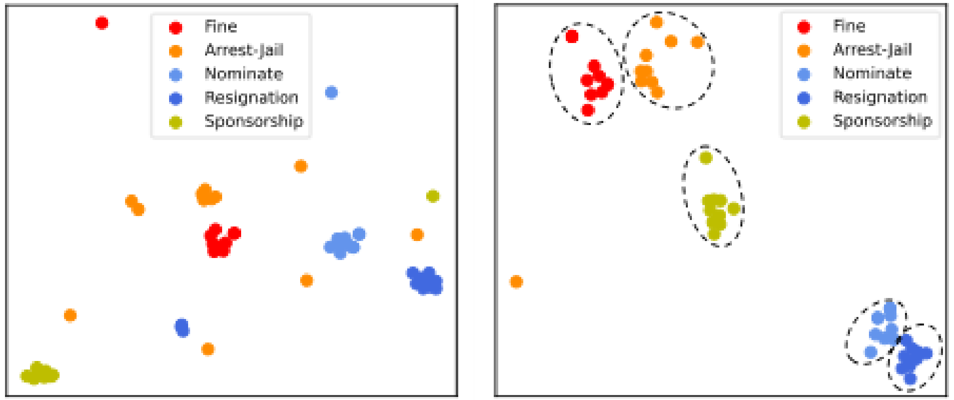
Introduction

$X^{(1)}$: Coach Santana *[sacked]* by Bahia. $Y^{(1)}$: {O, O, End-Position, O, O}
$X^{(2)}$: All but five were *[convicted]*. $Y^{(2)}$: {O, O, O, O, Convict}
...

Support Set

X^{q}: Tom Merritt *[fired]* from TwiT. Y^{q}: {O, O, End-Position, O, O}
...

Query Set



(a) PA-CRF (b) HCL-TAT

Figure 1: Visualization of triggers in the same episode on FewEvent test set. The left and right half shows support set representations without and with hybrid contrastive learning, respectively.

Approach

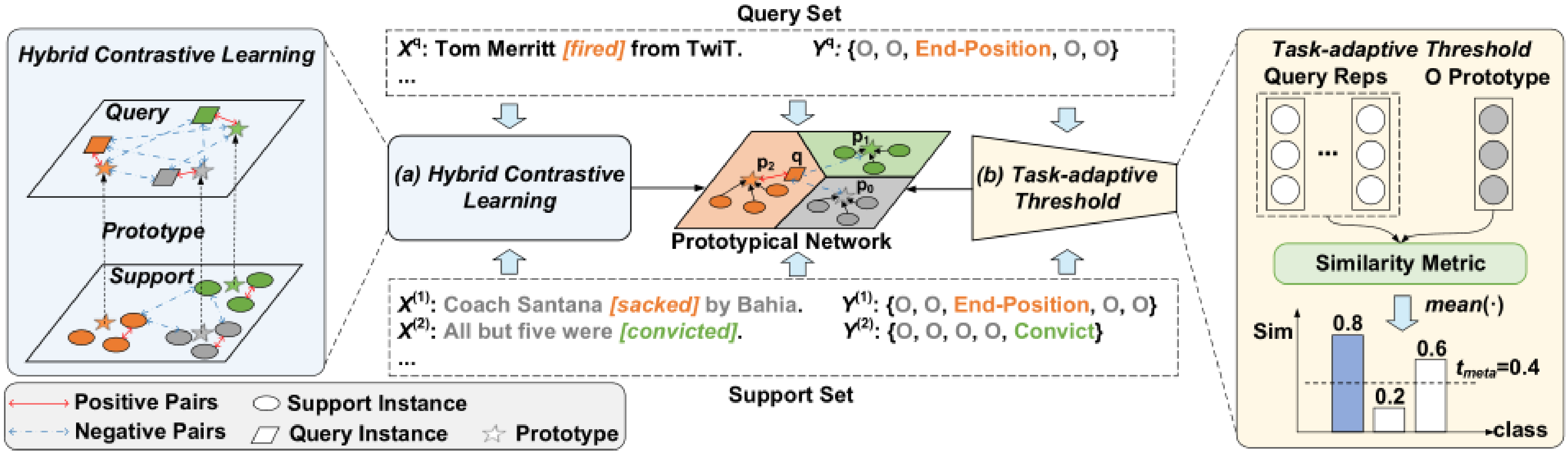
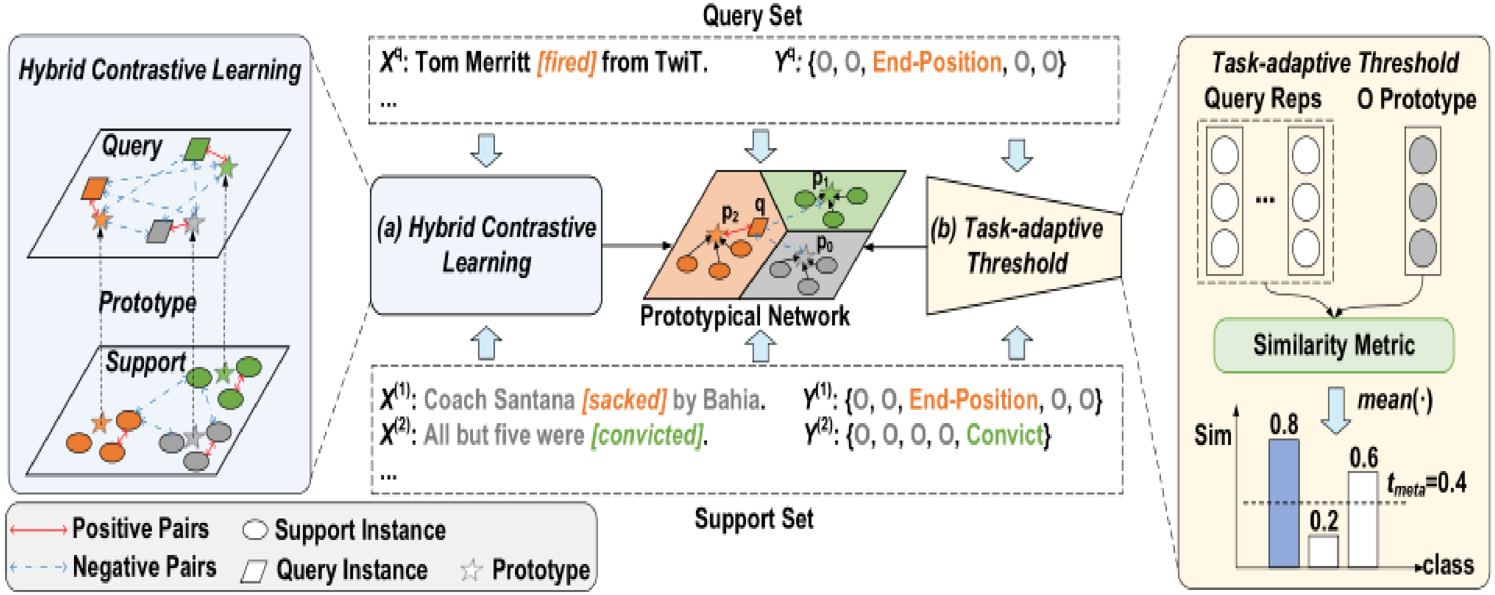


Figure 2: Overall framework of the proposed HCL-TAT model. HCL-TAT is based on a prototypical network, composed of two components: (a) hybrid contrastive learning including support-support contrastive learning and prototype-query contrastive learning; (b) task-adaptive threshold based on the logits in each episode.

Approach



$$\mathcal{X} = \{x_1, x_2, ..., x_n\}$$

$$\mathcal{Y} = \{y_1, y_2, ..., y_n\}$$

$$\mathcal{S} = \{\mathcal{X}^{(i)}, \mathcal{Y}^{(i)}\}_{i=1}^{N \times K}$$

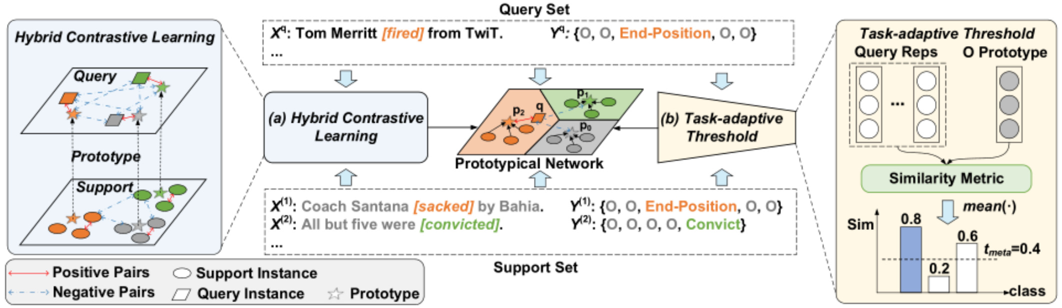
$$\mathcal{Q} = \{\mathcal{X}^{(i)}, \mathcal{Y}^{(i)}\}_{i=1}^{N \times M}$$

$$\mathcal{T} = \{\mathcal{S}, \mathcal{Q}\}$$

$$\mathcal{T}_{train} = \{\mathcal{T}_i\}_{i=1}^{M_{train}}$$

$$\mathcal{T}_{test} = \{\mathcal{T}_i\}_{i=1}^{M_{test}}$$

Approach



$$\{\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_n\} = f(\mathcal{X}, \theta), \tag{1}$$

$$\mathbf{p}_c = \frac{1}{K} \sum_{i \in \mathcal{S}(c)} \mathbf{h}_i, \; c = 0, 1, ..., N, \tag{2}$$

$$\mathcal{L}_{CE} = - \sum_{(x_i, y_i) \in \mathcal{Q}} \log P(y_i | x_i, \mathcal{S}), \tag{3}$$

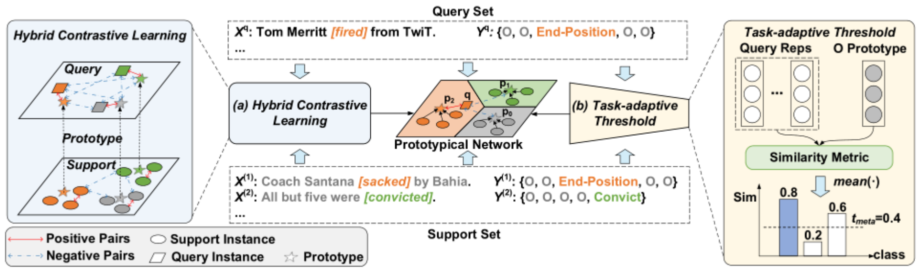
$$P(y_i | x_i, \mathcal{S}) = \frac{\exp(-d(\mathbf{h}_i, \mathbf{p}_{y_i}))}{\sum_{c \in \mathcal{C}} \exp(-d(\mathbf{h}_i, \mathbf{p}_c))}, \tag{4}$$

$$\frac{\partial \mathcal{L}_{CE}}{\partial \mathbf{h}_i} = \frac{\sum_n \Delta_n (\mathbf{p}^n - \mathbf{p}^{pos})}{1 + \sum_n \Delta_n}, \tag{5}$$

$$\frac{\partial \mathcal{L}_{CE}}{\partial \mathbf{p}^n} = \frac{\Delta_n \mathbf{h}_i}{1 + \sum_n \Delta_n}, \frac{\partial \mathcal{L}_{CE}}{\partial \mathbf{p}^{pos}} = - \frac{\sum_n \Delta_n \mathbf{h}_i}{1 + \sum_n \Delta_n}, \tag{6}$$

$$\Delta_n = \exp(\mathbf{h}_\mathrm{i} \cdot \mathbf{p}^\mathrm{n} \text{-} \mathbf{h}_\mathrm{i} \cdot \mathbf{p}^\mathrm{pos}) \,. \tag{7}$$

Approach

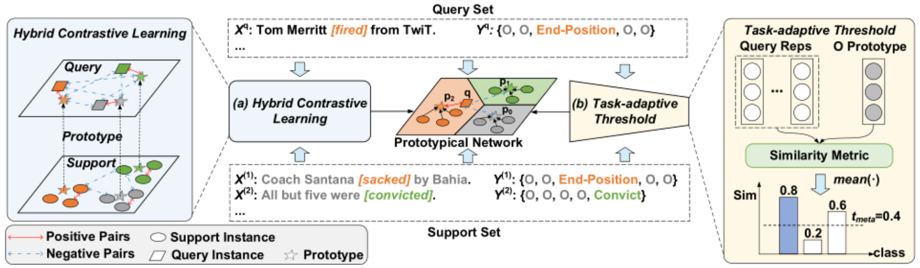


$$\tilde{\mathbf{h}}_i = \mathbf{W}_2\sigma(\mathbf{W}_1\mathbf{h}_i), \qquad (8)$$

$$\mathcal{L}_{SSCL} = \sum_{(x_i, y_i)\in\mathcal{S}} \mathcal{L}_{SSCL_i}, \qquad (9)$$

$$\mathcal{L}_{SSCL_i} = -\log \frac{\exp(\tilde{\mathbf{h}}_i \cdot \tilde{\mathbf{h}}_j/\tau)}{\sum_{k\neq i}\exp(\tilde{\mathbf{h}}_i \cdot \tilde{\mathbf{h}}_k/\tau)}, \qquad (10)$$

Approach



$$\mathcal{L}_{PQCL} = \sum_{c \in \mathcal{C}} \sum_{(x_i, y_i) \in \mathcal{Q}_c^{pos}} \mathcal{L}_{PQCL_c^i}, \tag{11}$$

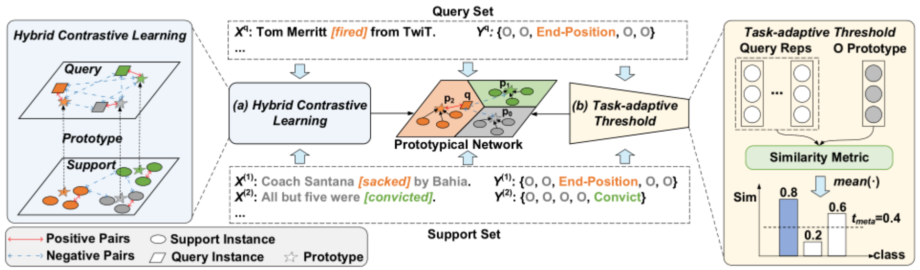
$$\mathcal{L}_{PQCL_c^i} = -\log \frac{sim_c^i}{sim_c^i + \sum_{(x_k, y_k) \in \mathcal{Q}_c^{neg}} sim_c^k}, \tag{12}$$

$$sim_c^i = \exp(\mathbf{p}_c \cdot \tilde{\mathbf{h}}_i / \tau). \tag{13}$$

$$t_{meta} = \frac{1}{|\mathcal{Q}|} \sum_{(x_i, y_i) \in \mathcal{Q}} P(y_i = 0 | x_i, \mathcal{S}). \tag{14}$$

$$\mathcal{L} = \mathcal{L}_{CE} + \alpha \mathcal{L}_{SSCL} + \beta \mathcal{L}_{PQCL}, \tag{15}$$

Approach



$$\begin{aligned}\mathcal{L}_{CE} &= -\log \frac{\exp(\mathbf{h}_i \cdot \mathbf{p}^{pos})}{\exp(\mathbf{h}_i \cdot \mathbf{p}^{pos}) + \sum_n \exp(\mathbf{h}_i \cdot \mathbf{p}^{n})} \\ &= -\log \frac{1}{1 + \sum_n \frac{\exp(\mathbf{h}_i \cdot \mathbf{p}^{n})}{\exp(\mathbf{h}_i \cdot \mathbf{p}^{pos})}} \\ &= \log(1 + \sum_n \frac{\exp(\mathbf{h}_i \cdot \mathbf{p}^{n})}{\exp(\mathbf{h}_i \cdot \mathbf{p}^{pos})}) \\ &= \log(1 + \sum_n \exp(\mathbf{h}_\mathrm{i} \cdot \mathbf{p}^\mathrm{n}\text{-}\mathbf{h}_\mathrm{i} \cdot \mathbf{p}^\mathrm{pos})),\end{aligned} \tag{16}$$

$$\begin{aligned}\frac{\partial \mathcal{L}_{CE}}{\partial \mathbf{h}_i} &= \frac{\sum_n \exp(\mathbf{h}_\mathrm{i} \cdot \mathbf{p}^\mathrm{n}\text{-}\mathbf{h}_\mathrm{i} \cdot \mathbf{p}^\mathrm{pos})\,|_{\mathbf{h}_i}}{1 + \sum_n \exp(\mathbf{h}_\mathrm{i} \cdot \mathbf{p}^\mathrm{n}\text{-}\mathbf{h}_\mathrm{i} \cdot \mathbf{p}^\mathrm{pos})} \\ &= \frac{\sum_n \exp(\mathbf{h}_\mathrm{i} \cdot \mathbf{p}^\mathrm{n}\text{-}\mathbf{h}_\mathrm{i} \cdot \mathbf{p}^\mathrm{pos})(\mathbf{p}^n - \mathbf{p}^{pos})}{1 + \sum_n \exp(\mathbf{h}_\mathrm{i} \cdot \mathbf{p}^\mathrm{n}\text{-}\mathbf{h}_\mathrm{i} \cdot \mathbf{p}^\mathrm{pos})} \\ &= \frac{\sum_n \Delta_n (\mathbf{p}^n - \mathbf{p}^{pos})}{1 + \sum_n \Delta_n}.\end{aligned} \tag{17}$$

$$\frac{\partial \mathcal{L}_{CE}}{\partial \mathbf{p}^n} = \frac{\Delta_n \mathbf{h}_i}{1 + \sum_n \Delta_n}, \tag{18}$$

$$\frac{\partial \mathcal{L}_{CE}}{\partial \mathbf{p}^{pos}} = -\frac{\sum_n \Delta_n \mathbf{h}_i}{1 + \sum_n \Delta_n}. \tag{19}$$

Experiments

Model	5-way-5-shot	5-way-10-shot	10-way-5-shot	10-way-10-shot
LoLoss	31.51 ± 1.56	31.70 ± 1.21	30.46 ± 1.38	30.32 ± 0.89
MatchLoss	30.44 ± 0.99	30.68 ± 0.78	28.97 ± 0.61	30.05 ± 0.93
DMBPN	37.51 ± 2.60	38.14 ± 2.32	34.21 ± 1.45	35.31 ± 1.69
Proto-dot†	41.54 ± 3.82	42.21 ± 0.68	33.27 ± 2.37	39.23 ± 2.95
Match†	30.09 ± 1.71	48.10 ± 1.38	28.94 ± 1.15	45.91 ± 1.98
Proto†	47.30 ± 2.55	54.81 ± 2.27	42.48 ± 1.00	50.14 ± 0.65
Vanilla CRF	59.01 ± 0.81	62.21 ± 1.94	56.00 ± 1.51	59.35 ± 1.09
CDT	59.30 ± 0.23	62.77 ± 0.12	56.41 ± 1.09	59.44 ± 1.83
PA-CRF	62.25 ± 1.42	64.45 ± 0.49	58.48 ± 0.68	61.64 ± 0.81
HCL-TAT	**66.96** ± 0.70	**68.80** ± 0.85	**64.19** ± 0.96	**66.00** ± 0.81

Table 1: F1 scores (10^{-2}) of evaluated methods on FewEvent test set. † means the model is re-implemented by ourselves. The best scores are highlighted in boldface, with $p < 0.02$ under t-test.

Experiments

Model	5-way-5-shot			5-way-10-shot		
	P	R	F1	P	R	F1
HCL-TAT	**62.63** $\pm$ 2.31	72.04 $\pm$ 1.93	**66.96** $\pm$ 0.70	**63.87** $\pm$ 2.35	74.65 $\pm$ 1.36	**68.80** $\pm$ 0.85
w/o SSCL	59.61 $\pm$ 2.48	71.65 $\pm$ 1.72	65.03 $\pm$ 0.82	60.22 $\pm$ 4.78	74.38 $\pm$ 1.81	66.42 $\pm$ 2.34
w/o PQCL	57.50 $\pm$ 1.80	71.88 $\pm$ 1.52	63.85 $\pm$ 0.67	60.88 $\pm$ 2.18	72.63 $\pm$ 1.23	66.21 $\pm$ 1.14
w/o HCL	49.52 $\pm$ 4.34	74.67 $\pm$ 3.36	59.38 $\pm$ 2.59	57.72 $\pm$ 2.72	73.35 $\pm$ 1.10	64.57 $\pm$ 1.69
w/o TAT	46.69 $\pm$ 1.25	**76.98** $\pm$ 0.29	58.12 $\pm$ 0.94	49.56 $\pm$ 1.11	76.33 $\pm$ 0.67	60.09 $\pm$ 0.92

Table 2: Precision, recall and F1 scores (10^{-2}) of ablation study results on FewEvent test set. When remove both HCL and TAT, the method degenerates to a Proto model.

Experiments

Model	FSTI	FSED
PA-CRF	63.68	62.25
HCL-TAT	68.18	66.96

Table 3: Average F1 scores (10^{-2}) of HCL-TAT and PA-CRF on FSTI and FSED tasks, on FewEvent test set under 5-way-5-shot setting.

Experiments

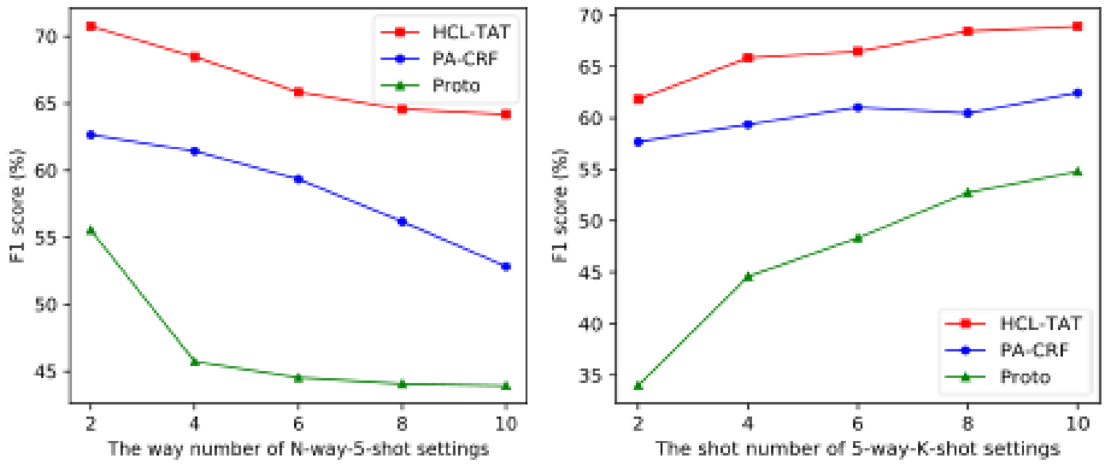
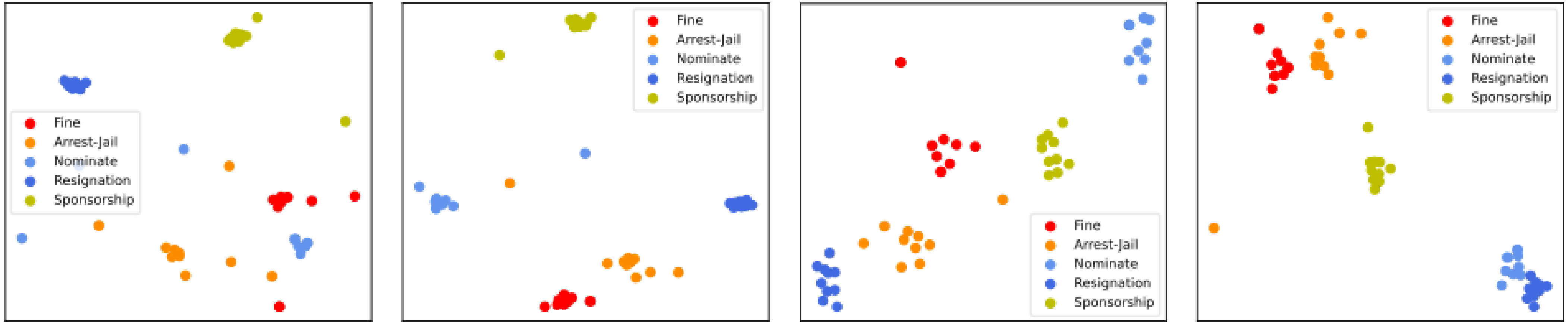


Figure 3: N-way-K-shot evaluations for three different models. The left part illustrates F1 scores in N-way-5-shot settings, and the right part illustrates F1 scores in 5-way-K-shot settings. We run each experiment once to analyze the tendency of F1 scores.

Experiments



(a) Proto (b) Proto+SSCL (c) Proto+PQCL (d) Proto+HCL

Figure 4: Visualization of trigger embeddings in the same episode on FewEvent test set, under 5-way-10-shot setting. From left to right, the visualazation results of four FSED models are given respectively.

Experiments

Subset	#Class	#Trigger	#Avg.Len
Train	80	69088	36.5
Valid	10	2274	38.6
Test	10	748	30.8

Table 4: The statistics of FewEvent Dataset. #Class, #Trigger and #Avg.Len denotes the number of classes, the number of triggers and the average length of sentences in each split part respectively.

Experiments

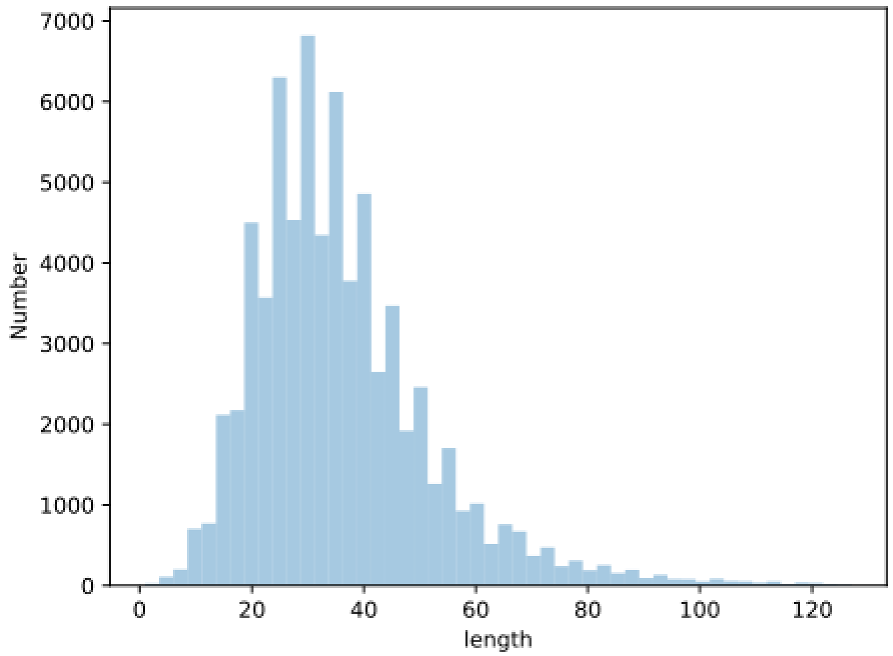


Figure 5: Length distribution of sentences in FewEvent dataset.

Experiments

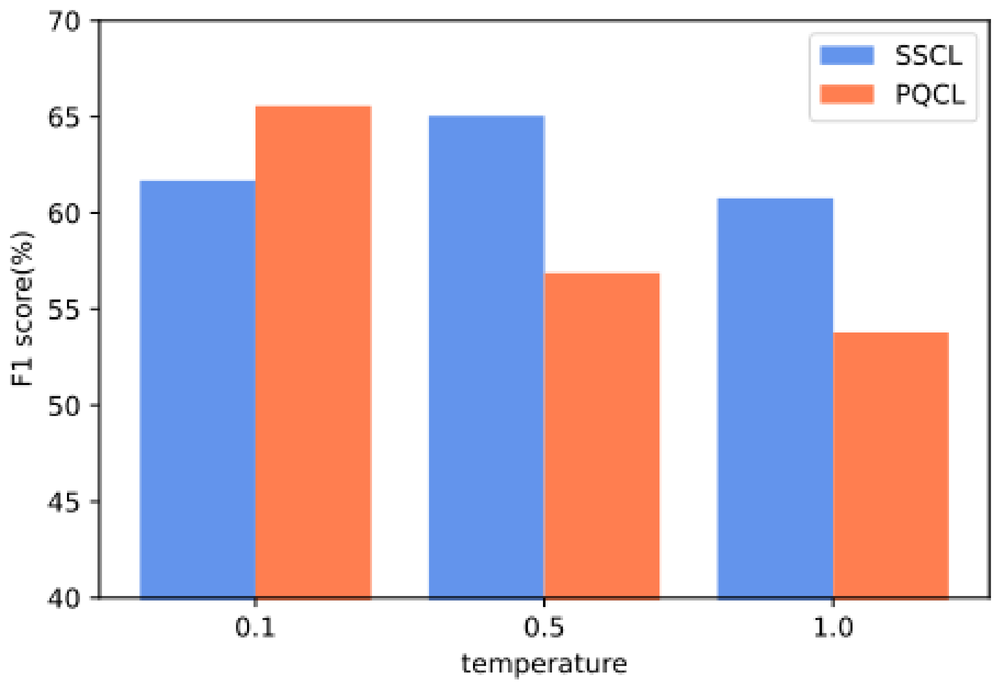


Figure 6: F1 scores(10^{-2}) over different temperature values on the two contrastive losses. The results are obtained under 5-way-5-shot setting in FewEvent test set.

Thank you！