Label-enhanced Prototypical Network with Contrastive Learning for Multi-label Few-shot Aspect Category Detection

Han Liu
Dalian University of Technology
Dalian, China
liu.han.dut@gmail.com

Feng Zhang
Peking University
Beijing, China
zhangfeng@stu.pku.edu.cn

Xiaotong Zhang*
Dalian University of Technology
Dalian, China
zxt.dut@hotmail.com

Siyang Zhao
Dalian University of Technology
Dalian, China
zhao_siyang@mail.dlut.edu.cn

Xianchao Zhang*
Dalian University of Technology
Dalian, China
xczhang@dlut.edu.cn

Hong Yu
Dalian University of Technology
Dalian, China
hongyu@dlut.edu.cn

Junjie Sun
Dalian University of Technology
Dalian, China
sunjunjiedlut@hotmail.com

—— KDD 2022

2022. 7. 13 • ChongQing

gesis
Leibniz-Institut
für Sozialwissenschaften

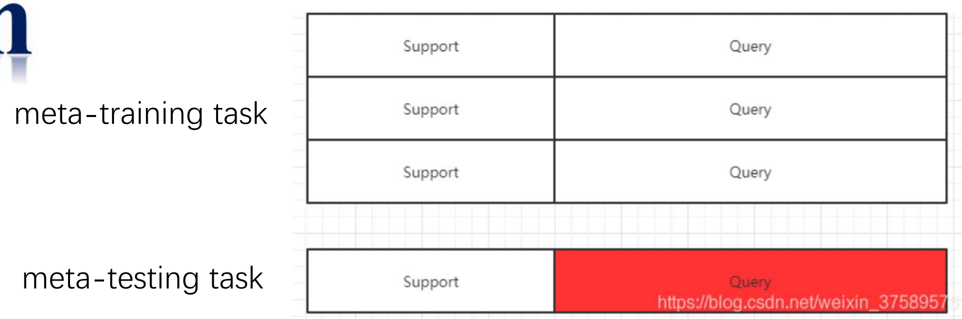
Reported by Sijin Liu

1.Introduction

2.Method

3.Experiments

Introduction



Support Set

staff	(1) It is the staff and food quality that really needs fixing. (2) The views are amazing from any location, staff is friendly and the food was great too!
food	(1) It is the staff and food quality that really needs fixing. (2) The views are amazing from any location, staff is friendly and the food was great too!
experience	(1) Incredible spa experience! (2) The food is always good and service has always been a great experience.

Query Set

experience and staff	(1) It was such a horrible experience, she was rude, unmannered and non professional great clip should not retain such a waste employee!
staff and food	(2) The pool is gorgeous, the rooms clean, delicious food, and staff that went above and beyond to help us enjoy our stay.
food	(3) We had breakfast the next morning on the first floor and the food was surprisingly good.

Figure 1: A meta-task example in 3-way 2-shot setting. The first column denotes the aspect label and the second column denotes the corresponding review sentence. As each review sentence may contain multiple aspects, we use different color background to mark the key words. The words in gray describe irrelevant (noisy) aspects, and the words in other colors represent the target aspects.

There exist several challenging issues:

(1) Simply calculating the prototype by averaging intra-class support samples may cause that different aspects share an identical prototype.

(2) When learning the prototype for a target aspect in multi-label scenarios,as each sentence probably contains several aspects, some irrelevant aspects will inevitably disturb the learning procedure.

(3) Due to the diversity of human expression, different sentences may contain different numbers of aspects, so it is urgently needed to design an effective model to automatically predict the number of aspects in a sentence.

Method

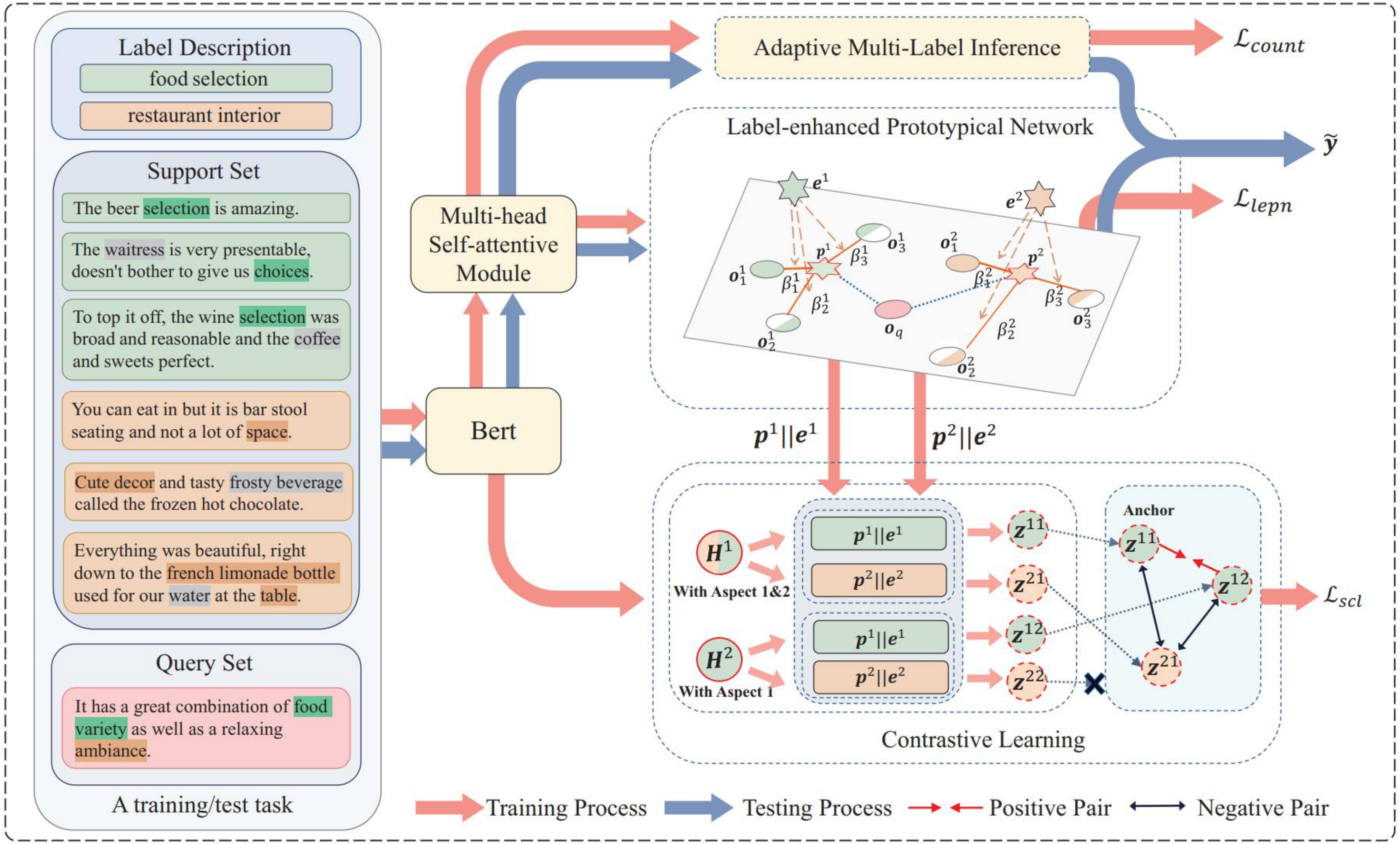
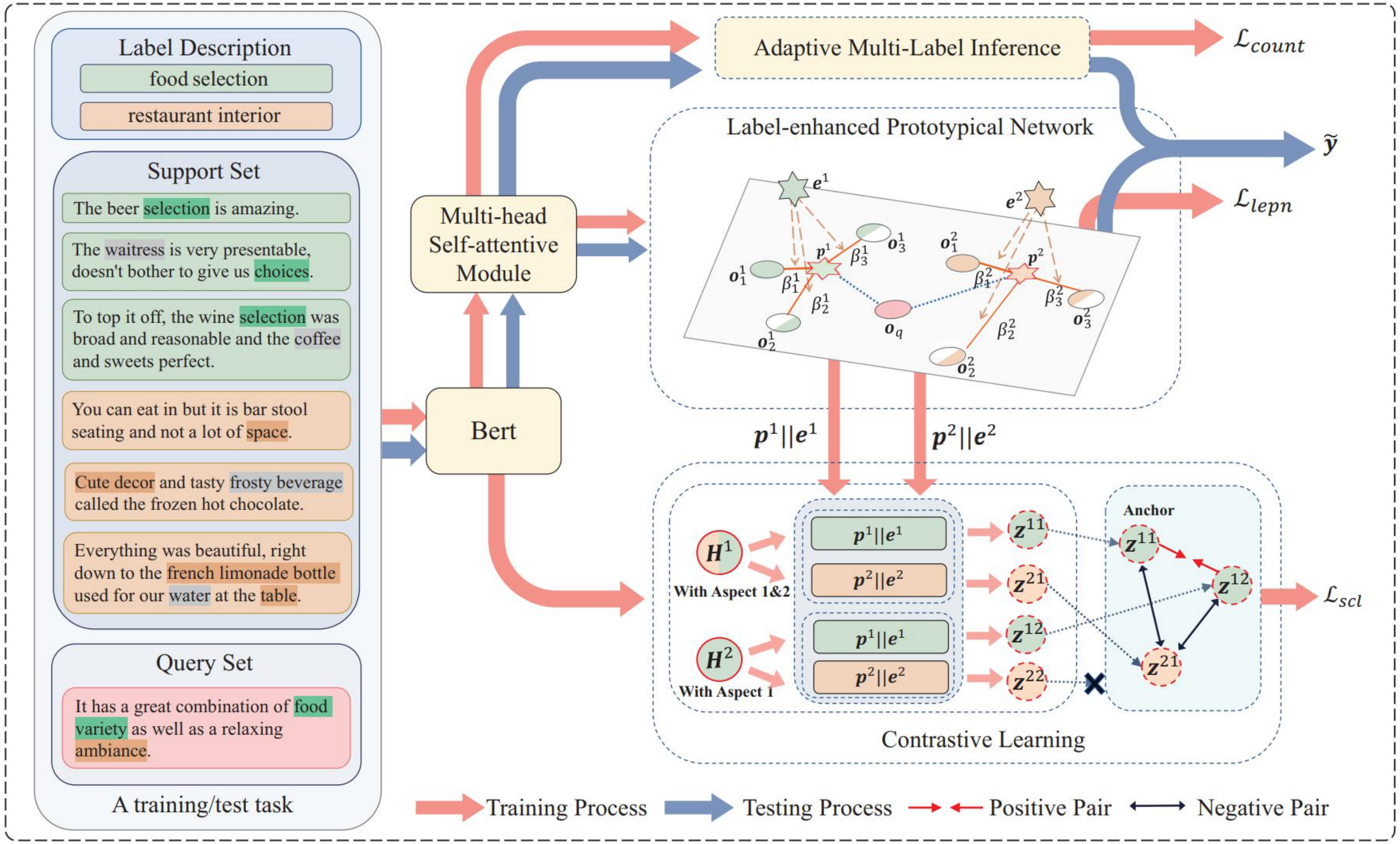


Figure 2: Illustration of our proposed method LPN.

Method



Feature Extraction

$$H = [h_1, h_2, ..., h_T] \in \mathbb{R}^{d\times T}$$

$$A = \text{softmax}\left(F_2 \tanh(F_1 H)\right), \quad (1)$$

$$M = HA^T, \quad (2)$$

$$o = F_3[m_1||m_2||...||m_R], \quad (3)$$

Label-enhanced Prototypical Network

Considering the N-way K-shot setting, we have a support set $\mathcal{S}$ which can be represented by $\mathcal{S} = \{x_1^1, x_2^1, ..., x_K^1, ..., x_1^N, x_2^N, ..., x_K^N\}$

$$O = \{o_1^1, o_2^1, ..., o_K^1, ..., o_1^N, o_2^N, ..., o_K^N\}$$

$$E = \{e^1, e^2, ..., e^N\}$$

$$\alpha_j^i = {o_j^i}^T W e^i, \quad (4)$$

$$\alpha_j^i = {o_j^i}^T U V^T e^i = \mathbf{1}^T (U^T o_j^i \circ V^T e^i), \quad (5)$$

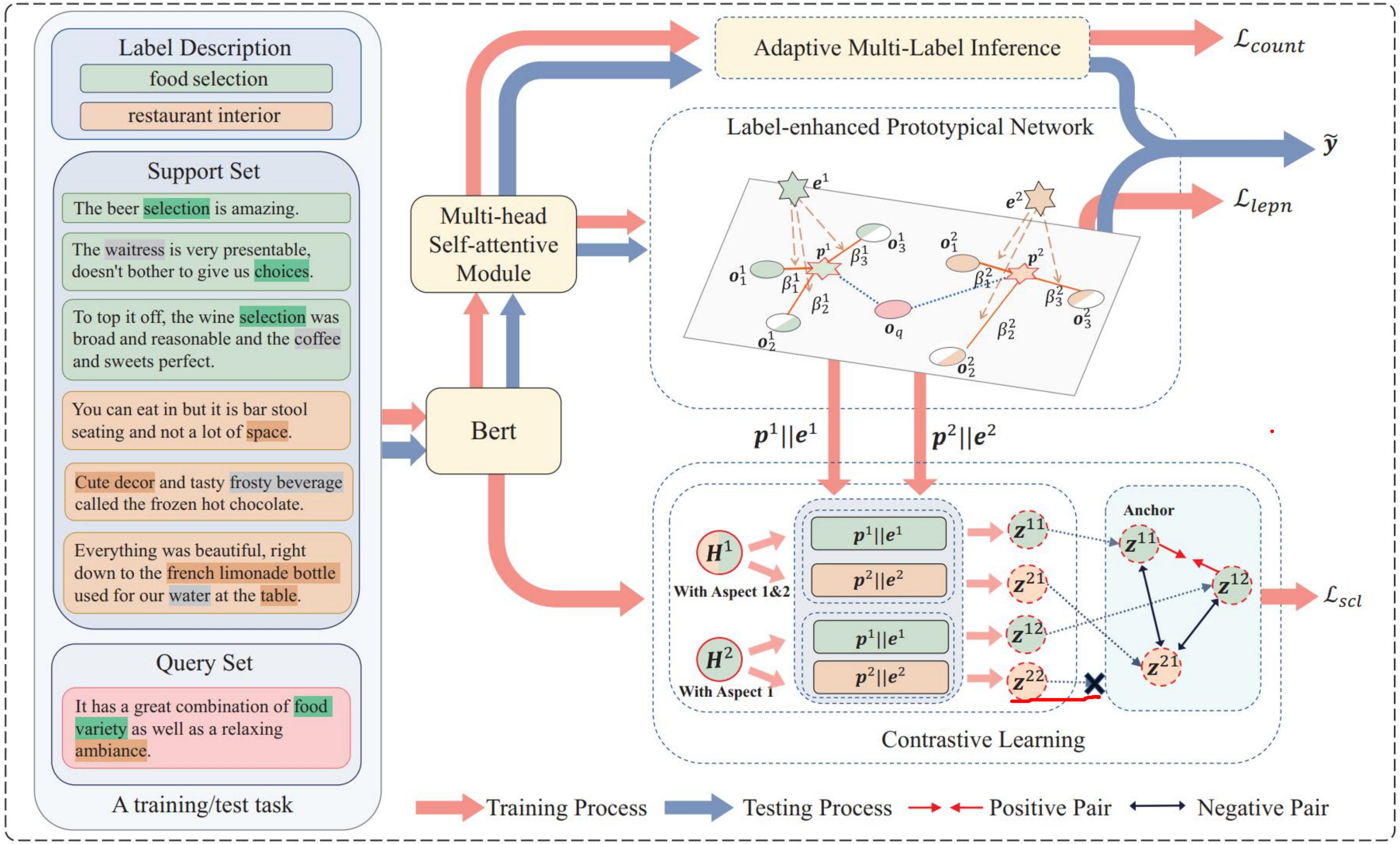
$$\beta_j^i = \frac{\exp(\alpha_j^i)}{\sum_{j'=1}^{K} \exp(\alpha_{j'}^i)}. \quad (6)$$

$$p^i = \sum_{j=1}^{K} \beta_j^i o_j^i. \quad (7)$$

$$p(y = y^i | x, \mathcal{S}) = \frac{\exp(-||o - p^i||_2^2)}{\sum_{j=1}^{N} \exp(-||o - p^j||_2^2)}, \quad (8)$$

$$\mathcal{L}_{lepn} = \frac{1}{|Q|} \sum_{x\in Q} \sum_{i=1}^{N} -y^i \log p(y = y^i | x, \mathcal{S}), \quad (9)$$

Method



$$I = \{z^{ij} \in Z | y^{ij} = 1\} \quad \Lambda^{ij} = \{z^{ik} \in \Gamma^{ij} | y^{ik} = y^{ij} = 1\} \quad \Gamma^{ij} = \{I \backslash z^{ij}\}$$

Integrating with Contrastive Learning

$$P = \{p^1, p^2, ..., p^N\}$$

$$E = \{e^1, e^2, ..., e^N\}$$

$$\{a^1, a^2, ..., a^N\}, \text{ where } a^i = [p^i || e^i] \in \mathbb{R}^{2d}$$

$$z^i = g^i H^T, \tag{10}$$

$$g^i = \text{softmax}((W_a a^i + b_a)^T H), \tag{11}$$

$$Z = \{z^{ij} \in \mathbb{R}^d | i \in \{1, 2, ..., N\}, j \in \{1, 2, ..., N_t\}\}$$

$$Y = \{y^{ij} \in \{0, 1\} | i \in \{1, 2, ..., N\}, j \in \{1, 2, ..., N_t\}\}$$

$$\mathcal{L}_{scl}^{ij} = -\frac{1}{|\Lambda^{ij}|} \sum_{z^{ik} \in \Lambda^{ij}} \log \frac{\exp(z^{ij} \cdot z^{ik} / \tau)}{\sum_{z^* \in \Gamma^{ij}} \exp(z^{ij} \cdot z^* / \tau)}, \tag{12}$$

$$\mathcal{L}_{scl} = \frac{1}{|I|} \sum_{z^{ij} \in I} \mathcal{L}_{scl}^{ij}. \tag{13}$$

Adaptive Multi-label Inference

$$n_l = \text{softmax}(W_l o + b_l), \tag{15}$$

$$\mathcal{L}_{count} = \frac{1}{|S \cup Q|} \sum_{x \in S \cup Q} -\mathbf{1}^T (t_l \circ \log(n_l)), \tag{16}$$

$$\mathcal{L}_{total} = \mathcal{L}_{lepn} + \gamma \mathcal{L}_{scl} + \lambda \mathcal{L}_{count}, \tag{17}$$

Experiments

Table 2: Dataset statistics. #Aspects and #Sentences denote the number of aspects and sentences respectively.

Dataset	Split	#Aspects	#Sentences
FewAsp (single)	Training	64	12800
	Validation	16	3200
	Testing	20	4000
FewAsp (multi)	Training	64	25600
	Validation	16	6400
	Testing	20	8000
FewAsp	Training	64	40320
	Validation	16	10080
	Testing	20	12600

Table 3: Hyperparameters of our proposed method LPN.

Model	d	d'	R	k	λ	γ	τ
LPN	768	256	4	100	0.1	0.01	0.1

Experiments

Table 4: Average AUC and macro-F1 score on FewAsp (single).

Model	5-way 5-shot		5-way 10-shot		10-way 5-shot		10-way 10-shot	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1
Matching Network [35]	97.05	81.89	97.49	84.62	96.30	70.95	96.72	73.28
Prototypical Network [30]	96.49	83.30	97.53	86.29	95.97	74.23	96.71	76.83
Relation Network [32]	93.31	75.79	90.86	72.02	91.81	63.78	90.54	61.15
Graph Network [28]	96.54	81.45	97.46	85.04	95.45	70.75	96.97	77.84
IMP [2]	96.65	83.69	97.47	86.14	96.00	73.80	96.91	77.09
Proto-HATT [9]	96.45	83.33	97.62	86.71	95.71	73.42	97.00	77.65
Proto-AWATT [16]	97.56	86.71	97.96	88.54	97.01	80.28	97.55	82.97
LPN (o, o)	97.88	87.62	98.48	90.31	98.13	83.99	98.53	85.95
LPN (w, o)	99.22	92.61	99.35	93.57	99.11	89.35	**99.32**	**91.08**
LPN (w, w)	**99.29**	**94.43**	**99.49**	**94.40**	**99.14**	**89.40**	99.28	90.43

Experiments

Table 5: Average AUC and macro-F1 score on FewAsp (multi).

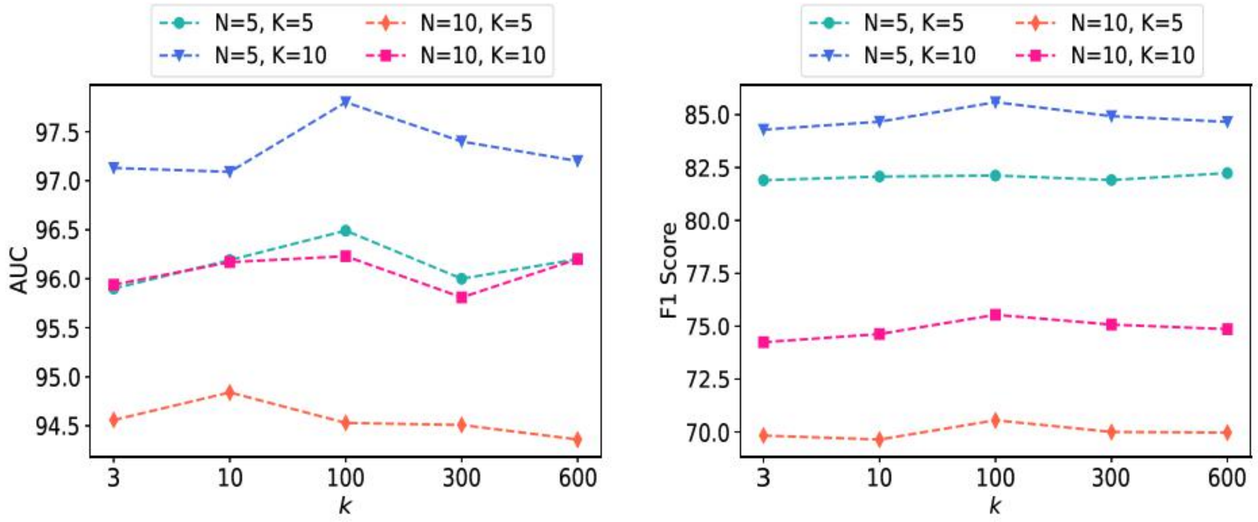
Model	5-way 5-shot		5-way 10-shot		10-way 5-shot		10-way 10-shot	
	AUC	F1	AUC	F1	AUC	F1	AUC	F1
Matching Network [35]	89.54	65.70	91.38	69.02	88.28	50.86	89.94	54.42
Prototypical Network [30]	89.67	67.88	91.60	72.32	88.01	52.72	90.68	58.92
Relation Network [32]	84.91	58.38	86.21	61.37	84.22	43.71	84.72	44.85
Graph Network [28]	87.97	59.25	90.45	64.63	86.05	45.42	88.44	48.49
IMP [2]	90.12	68.86	92.29	73.51	88.71	53.96	91.10	59.86
Proto-HATT [9]	91.10	69.15	93.03	73.91	90.44	55.34	92.38	60.21
Proto-AWATT [16]	91.45	71.72	93.89	77.19	89.80	58.89	92.34	66.76
LPN (o, o)	93.09	72.45	94.92	76.89	92.95	61.33	94.62	66.39
LPN (w, o)	95.43	78.82	96.22	81.70	94.29	66.36	95.43	71.08
LPN (w, w)	**95.66**	**79.48**	**96.55**	**82.81**	**94.51**	**67.28**	**95.66**	**71.87**

Experiments

Table 6: Average AUC and macro-F1 score on FewAsp.

Model	5-way 5-shot AUC	5-way 5-shot F1	5-way 10-shot AUC	5-way 10-shot F1	10-way 5-shot AUC	10-way 5-shot F1	10-way 10-shot AUC	10-way 10-shot F1
Matching Network [35]	90.76	67.14	92.39	70.09	88.44	51.27	89.90	54.61
Prototypical Network [30]	88.88	66.96	91.77	73.27	87.35	52.06	90.13	59.03
Relation Network [32]	85.56	59.52	86.98	62.78	84.94	45.62	83.77	44.70
Graph Network [28]	89.48	61.49	92.35	69.89	87.35	47.91	90.19	56.06
IMP [2]	89.95	68.96	92.30	74.13	88.50	54.14	90.81	59.84
Proto-HATT [9]	91.54	70.26	93.43	75.24	90.63	57.26	92.86	61.51
Proto-AWATT [16]	93.35	75.37	95.28	80.16	92.06	65.65	93.42	69.70
LPN (o, o)	94.15	76.19	95.85	80.37	94.03	65.72	94.98	69.22
LPN (w, o)	96.41	**82.26**	**97.43**	**85.81**	95.26	71.25	96.23	75.49
LPN (w, w)	**96.45**	82.22	97.15	84.90	**95.36**	**71.42**	**96.55**	**76.51**

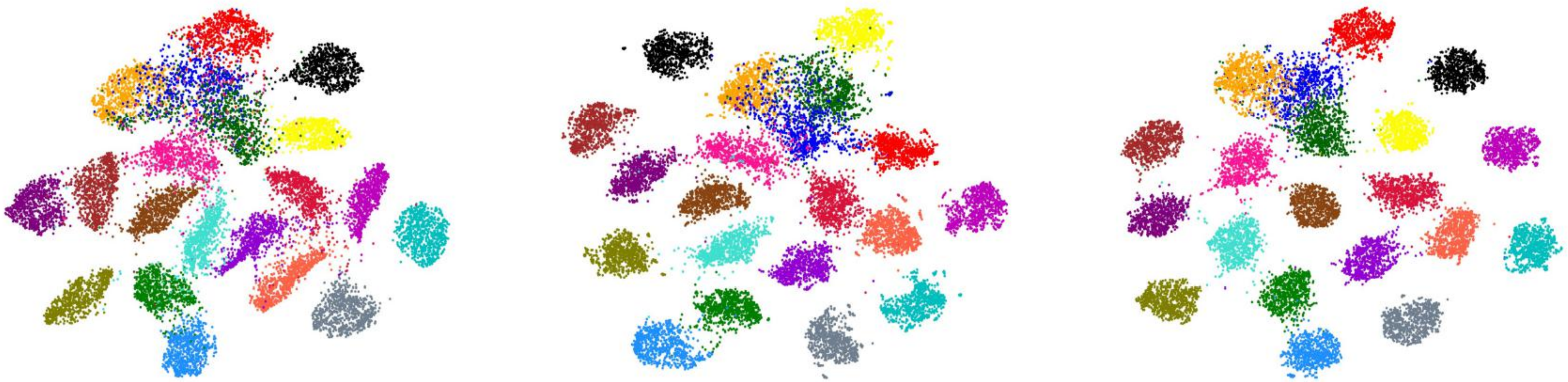
Experiments



(a) AUC (b) F1 Score

Figure 3: The performance of LPN (w,w) with different k values on the validation set of FewAsp.

Experiments



(a) Pic.1: LPN (o, o). (b) Pic.2: LPN (w, o). (c) Pic.3: LPN (w, w).

Figure 4: Visualization of prototype embeddings obtained from LPN (o, o), LPN (w, o) and LPN (w, w) respectively.