Chongqing University of Technology

Label-Driven Denoising Framework for Multi-Label Few-Shot Aspect Category Detection

Fei Zhao[1*] Yuchen Shen[2*] Zhen Wu[1†] Xinyu Dai[1]
[1]National Key Laboratory for Novel Software Technology, Nanjing University
[2]School of Information and Software Engineering,
University of Electronic Science and Technology of China
zhaof@smail.nju.edu.cn, alexchicharitoshen@gmail.com
{wuz, daixinyu}@nju.edu.cn

https://github.com/1429904852/LDF

—— EMNLP 2022

Reported by Yuyang Lai

1.Introduction

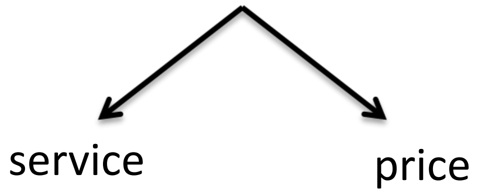
2.Method

3.Experiments

Introduction

Aspect Category Detection (ACD)

The service is good although rooms are pretty expensive.



FS-ACD

Support set	
Aspect Category	Sentences
(A) food_food_meat_burger	(1) *first time, burger was not fully cooked and my smash fries were cold.* (2) *food was over priced, but okay not great.*
(B) food_mealtype_lunch	(1) *my brother and i stopped in for lunch.* (2) *lunch has a great option of picking one or two food with rice.*
(C) restaurant_location	(1) *i prefer the other location to be honest.* (2) *there's a new standard in town.*
Query set	
Aspect Category	Sentences
(B)	(1) *went back today for lunch.*
(A) and (C)	(2) *food is whats to be expected at a neighborhood grill.*

Table 1: An example of 3-way 2-shot meta-task. A sentence (instance) may belong to multiple aspects.

Introduction

- due to lack of sufficient supervised data, the previous methods easily catch noisy words irrelevant to the current aspect category, which largely affects the quality of the generated prototype;

- the semantically-close aspect categories usually generate similar prototypes, which arenmutually noisy and confuse the classifier seriously

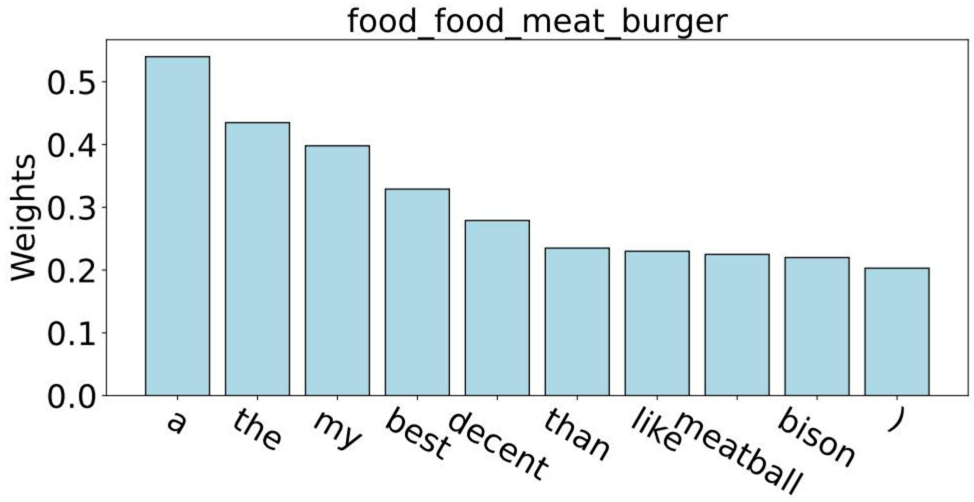


Figure 1: Visualization of the top-10 words for the prototype of aspect category *food_food_meat_burger* according to the attention weights of *Proto-AWATT*.

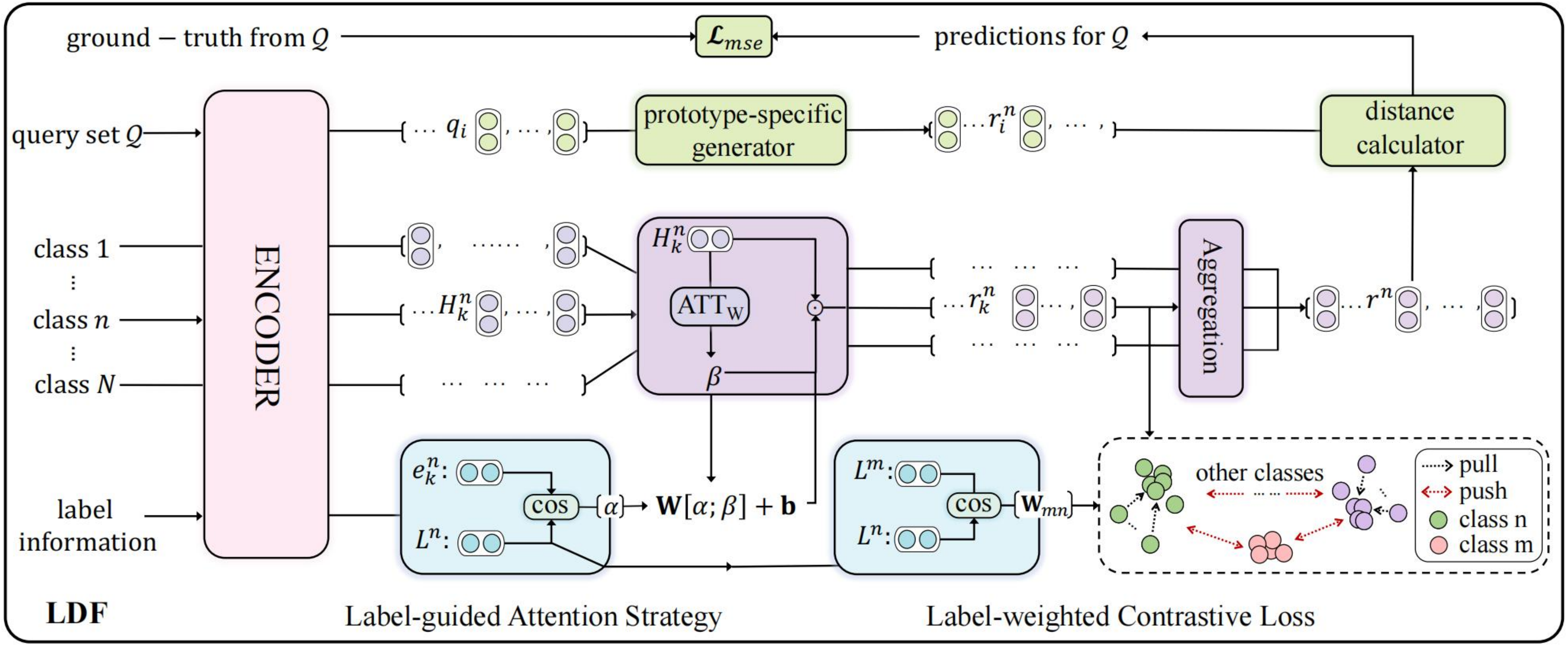
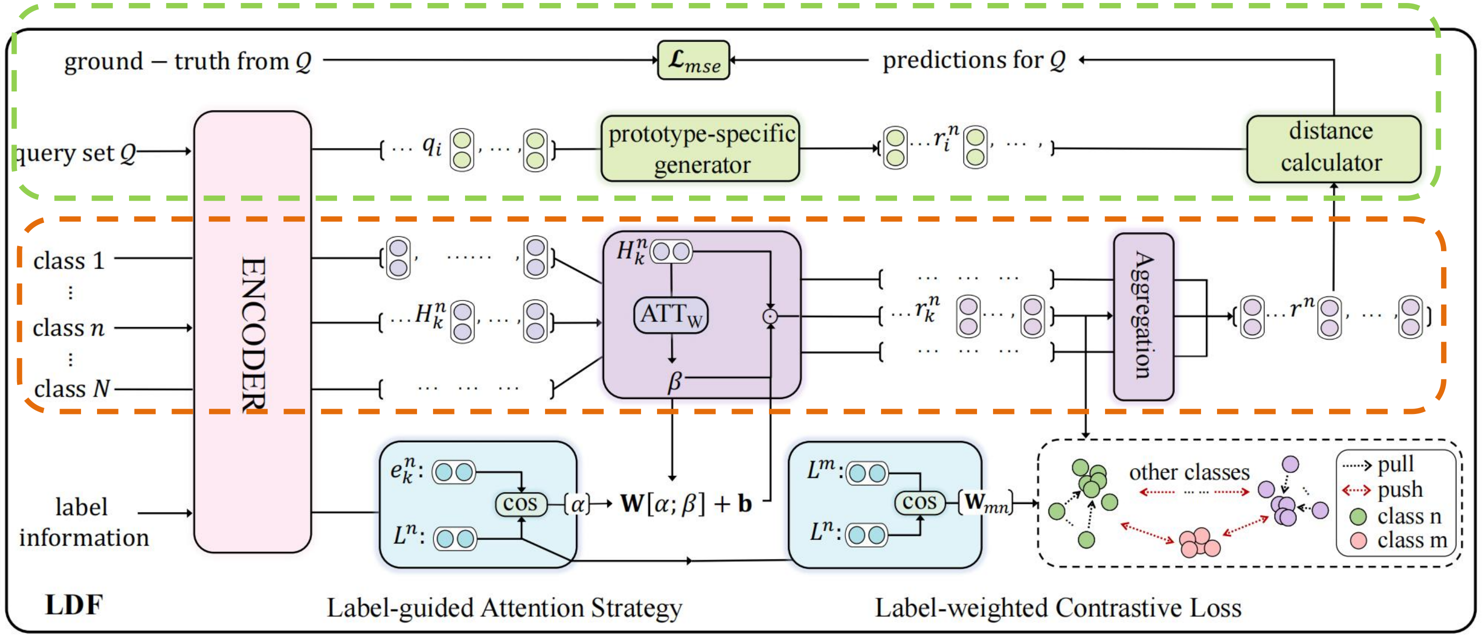


Figure 2: The overview of our proposed LDF framework.

Method



$$\mathcal{S} = \{s_k^n \mid k = 1, ..., K\}_{n=1}^{N}$$

$$\mathcal{Q} = \{(q_i, y_i) \mid y_i \in \mathbb{R}^N\}_{i=1}^{M}$$

$$s_k^n = \{w_1, w_2, ..., w_l\}$$

$$e_k^n = \{e_1, e_2, ..., e_l\}$$

$$\beta = \text{ATT}_W(H_k^n), \quad (1)$$

$$r_k^n = \beta H_k^n, \quad (2)$$

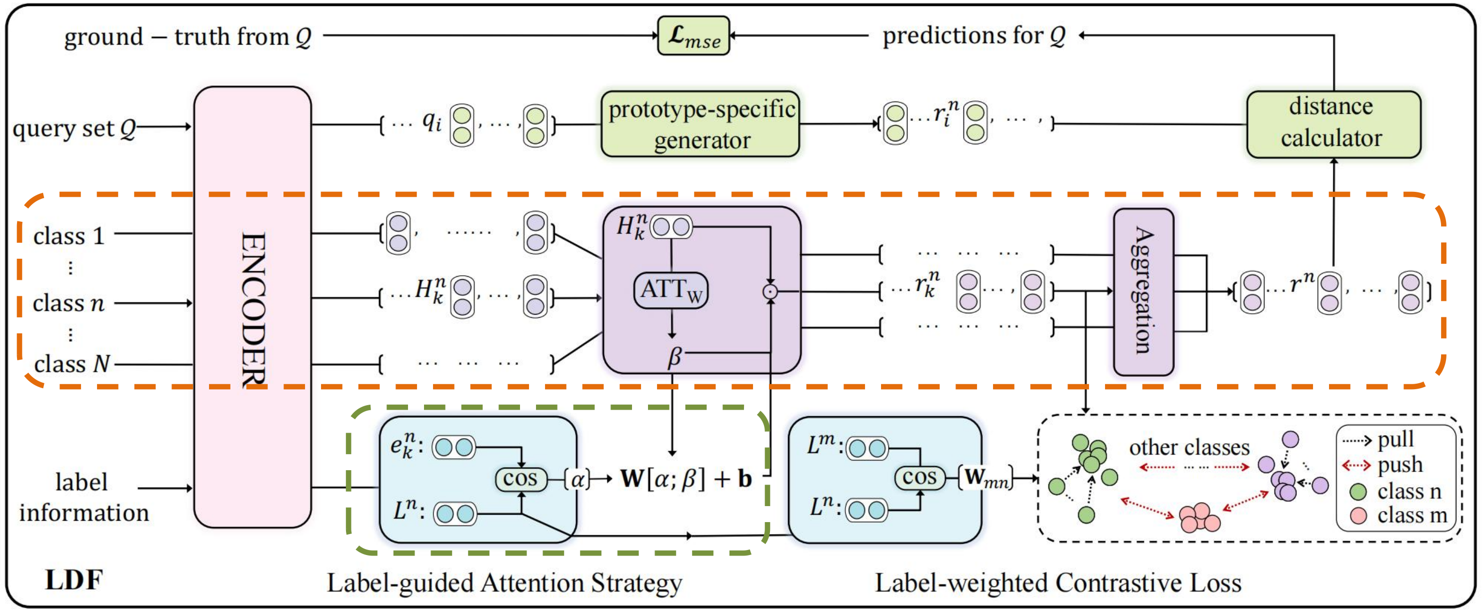
$$r^n = \text{Aggregation}(r_1^n, ..., r_K^n), \quad (3)$$

$$\{r^1, r^2, ..., r^n, ..., r^N\}$$

$$\hat{y}_i = \text{softmax}(-\text{ED}(r^n, r_i^n)), n \in [1, N] \quad (4)$$

$$\mathcal{L}_{mse} = \sum_{i=1}^{M} (\hat{y}_i - y_i)^2 \quad (5)$$

Method



$$\beta = \mathrm{ATT}_W(H_k^n), \quad (1)$$

$$r_k^n = \beta H_k^n, \quad (2)$$

$$r^n = \mathrm{Aggregation}(r_1^n, ..., r_K^n), \quad (3)$$

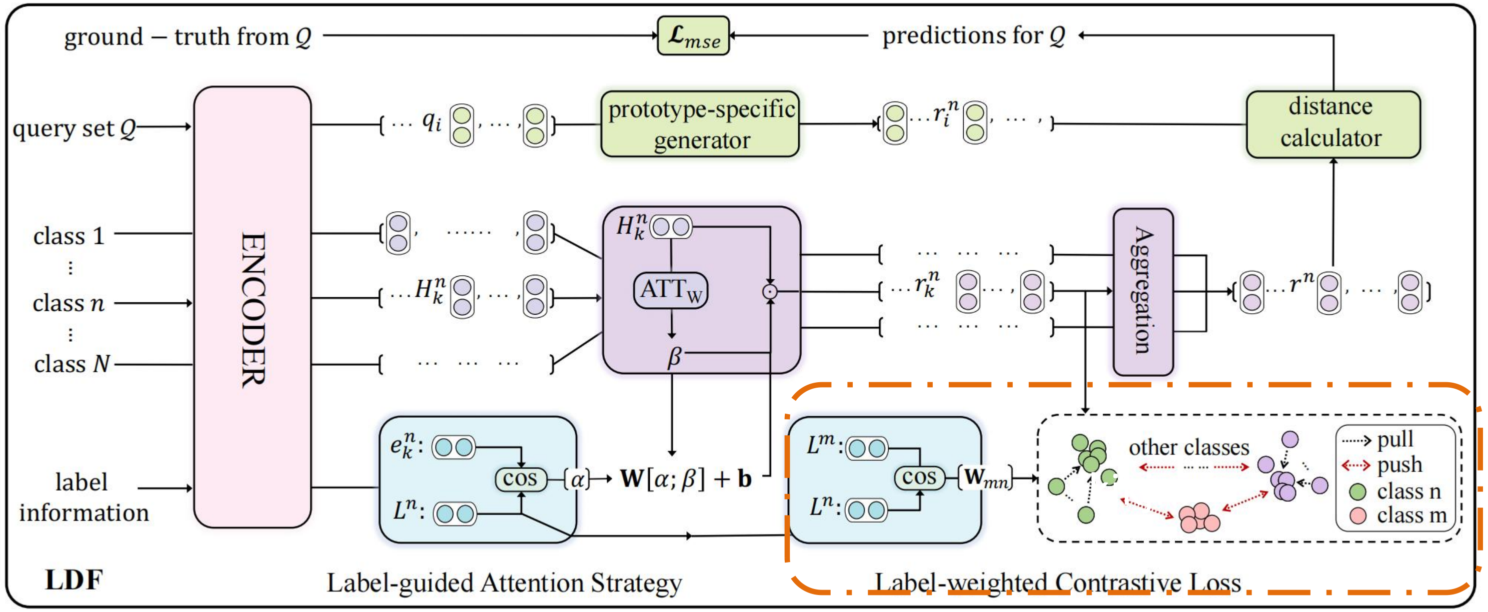
$$\{r^1, r^2, ..., r^n, ..., r^N\}$$

$$\alpha = \cos(L^n, e_k^n), \quad (6)$$

$$\theta = W_g[\alpha; \beta] + b_g \quad (7)$$

$$\widetilde{\theta} = \mathrm{softmax}(\theta) \quad (8)$$

Method



"food_food_meat_burger"

"food_mealtype_lunch"

"room_bed"

$$\mathcal{L}_{scl} = \sum_{(n,k)\in(N,K)} \frac{-1}{|P(n,k)|} \sum_{r_p^n \in P(n,k)} \log \frac{\exp(r_k^n \cdot r_p^n/\tau)}{\sum_{r_k^m \in (N,K)\setminus(n,k)} \exp(r_k^n \cdot r_k^m/\tau)} \quad (9)$$

$$w_{mn} = \cos(L^m, L^n), \quad (11)$$

$$\mathcal{L}_{lcl} = \sum_{(n,k)\in(N,K)} \frac{-1}{|P(n,k)|} \sum_{r_p^n \in P(n,k)} \log \frac{\exp(r_k^n \cdot r_p^n/\tau)}{\sum_{r_k^m \in (N,K)\setminus(n,k)} w_{mn} \cdot \exp(r_k^n \cdot r_k^m/\tau)} \quad (10)$$

$$\mathcal{L} = \mathcal{L}_{mse} + \lambda \mathcal{L}_{lcl} \quad (12)$$

Experiments

Dataset	#cls.	#inst./cls.	#inst.
FewAsp(single)	100	200	20000
FewAsp(multi)	100	400	40000
FewAsp	100	630	63000

Table 2: Statistics of three datasets. **#cls.** is the number of classes. **#inst.** is the total number of instances. **#inst./cls.** is the number of instances per class.

Experiments

Models	5-way 5-shot F1	5-way 5-shot AUC	5-way 10-shot F1	5-way 10-shot AUC	10-way 5-shot F1	10-way 5-shot AUC	10-way 10-shot F1	10-way 10-shot AUC
FewAsp								
Proto-HATT	70.26	91.54	75.24	93.43	57.26	90.63	61.51	92.86
LDF-HATT	**73.56**† ±0.47	**92.60**† ±0.23	**78.81**† ±0.93	**94.75**† ±0.43	**60.68**† ±0.92	**91.22** ±0.53	**67.13**† ±0.94	**94.12**† ±0.29
Δ	+3.30	+1.06	+3.57	+1.32	+3.42	+0.59	+5.62	+1.26
Proto-AWATT	75.37	93.35	80.16	95.28	65.65	92.06	69.70	93.42
LDF-AWATT	**78.27**† ±0.89	**94.65**† ±0.41	**81.87**† ±0.48	**95.71** ±0.26	**67.13**† ±0.41	**92.74** ±0.12	**71.97**† ±0.49	**94.29** ±0.25
Δ	+2.90	+1.30	+1.71	+0.43	+1.48	+0.68	+2.27	+0.87
FewAsp(single)								
Proto-HATT	83.33	96.45	86.71	97.62	73.42	95.71	77.65	97.00
LDF-HATT	**84.41**† ±0.46	**97.06** ±0.16	**88.15**† ±1.00	**98.12** ±0.31	**76.27**† ±1.08	**96.38** ±0.37	**80.54**† ±0.97	**97.45** ±0.14
Δ	+1.08	+0.61	+1.44	+0.50	+2.85	+0.67	+2.89	+0.45
Proto-AWATT	86.71	97.56	88.54	97.96	80.28	97.01	82.97	97.55
LDF-AWATT	**88.16**† ±0.62	**98.29** ±0.32	**89.32** ±0.92	**98.38** ±0.13	**81.73**† ±0.96	**97.51** ±0.33	**84.20**† ±0.21	**97.96** ±0.30
Δ	+1.45	+0.73	+0.78	+0.42	+1.45	+0.50	+1.23	+0.41
FewAsp(multi)								
Proto-HATT	69.15	91.10	73.91	93.03	55.34	90.44	60.21	92.38
LDF-HATT	**72.13**† ±0.79	**92.19**† ±0.33	**76.52**† ±0.74	**93.68** ±0.36	**59.10**† ±1.04	**91.00** ±0.51	**65.31**† ±0.57	**92.99** ±0.24
Δ	+2.98	+1.09	+2.61	+0.65	+3.76	+0.56	+5.10	+0.61
Proto-AWATT	71.72	91.45	77.19	93.89	58.89	89.80	66.76	92.34
LDF-AWATT	**73.38**† ±0.73	**92.62**† ±0.32	**78.81**† ±0.19	**94.34** ±0.15	**62.06**† ±0.54	**90.87**† ±0.48	**68.23**† ±0.98	**92.93** ±0.44
Δ	+1.66	+1.17	+1.62	+0.44	+3.17	+1.07	+1.47	+0.59

Experiments

Models	5-way 5-shot F1	5-way 5-shot AUC	5-way 10-shot F1	5-way 10-shot AUC	10-way 5-shot F1	10-way 5-shot AUC	10-way 10-shot F1	10-way 10-shot AUC
Proto-AWATT	75.37	93.35	80.16	95.28	65.65	92.06	69.70	93.42
Proto-AWATT+LAS	77.31±1.96	94.42±0.67	81.19±0.84	95.49±0.36	66.48±3.02	92.54±0.70	71.12±1.14	94.26±0.40
Proto-AWATT+LCL	77.06±0.71	94.20±0.26	80.78±0.39	95.44±0.22	66.20±1.26	92.38±0.45	70.83±0.66	94.07±0.33
Proto-AWATT+SCL	76.11±1.76	93.67±0.80	80.24±2.99	95.31±1.01	65.76±2.17	92.36±0.60	70.03±2.69	93.93±0.67
LDF-AWATT	**78.27**±0.89	**94.65**±0.41	**81.87**±0.48	**95.71**±0.26	**67.13**±0.41	**92.74**±0.12	**71.97**±0.49	**94.29**±0.25

Table 4: Ablation study over two main components on FewAsp dataset. The ablation results of FewAsp(single) and FewAsp(multi) datasets are included in **Appendix** A.3.

Experiments

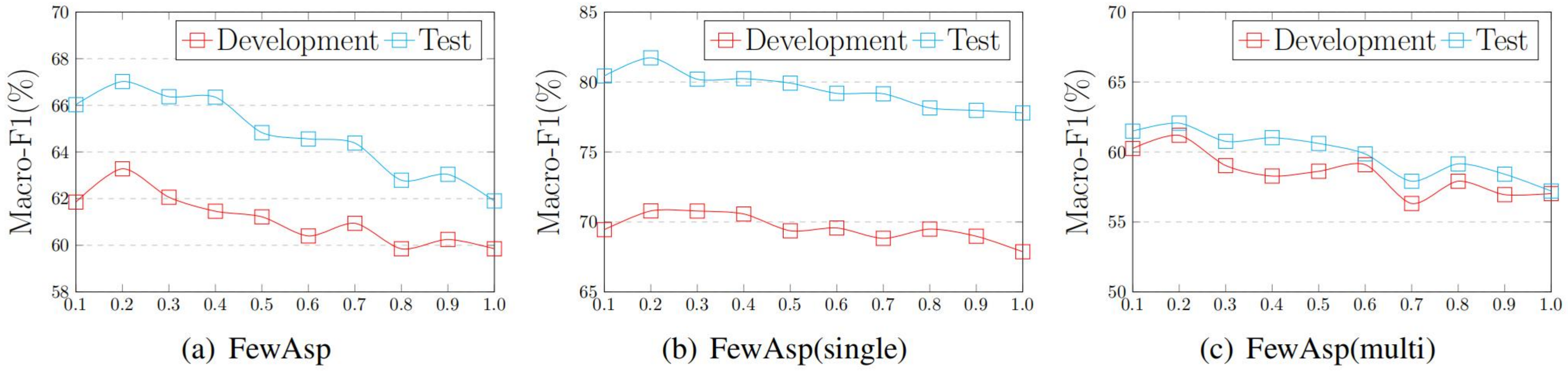


Figure 3: Effect of λ in the 10-way 5-shot setting on three dataset.

Experiments

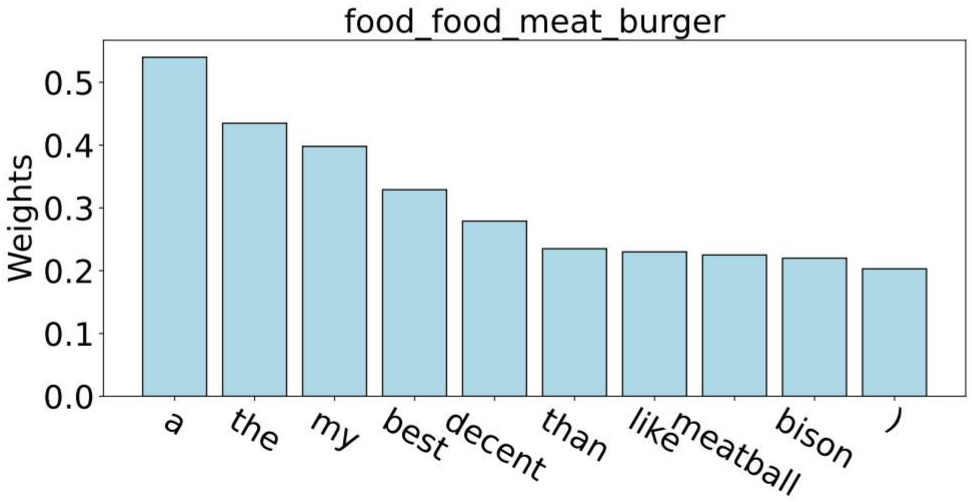


Figure 1: Visualization of the top-10 words for the prototype of aspect category *food_food_meat_burger* according to the attention weights of *Proto-AWATT*.

Figure 4: Visualize the top-10 words for the prototype of aspect category *food_food_meat_burger* based on the attention weights of *Proto-AWATT+LAS*.

Experiments

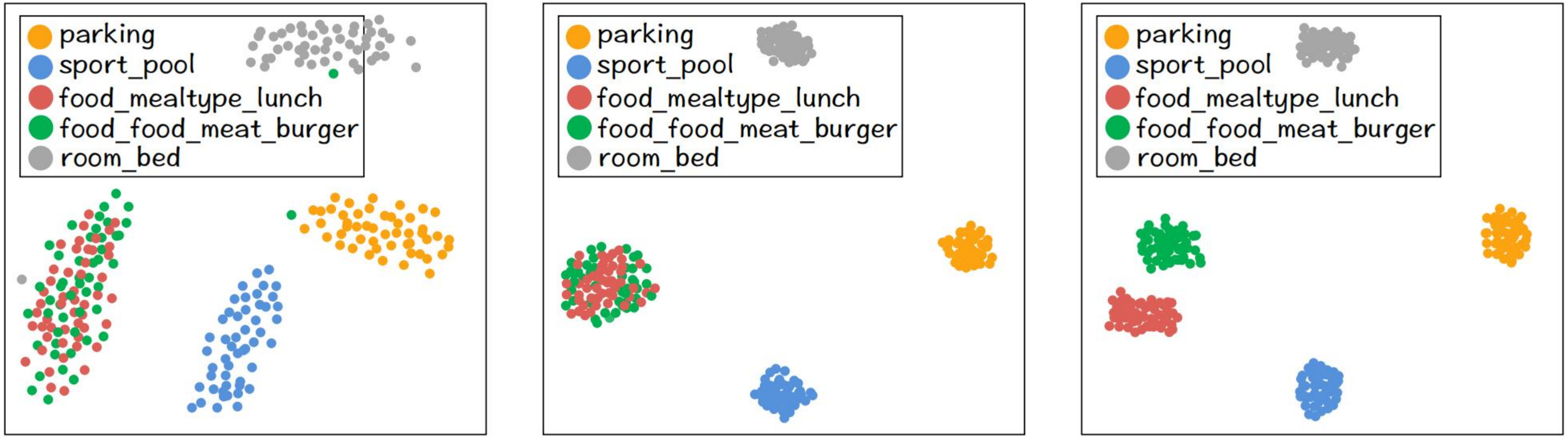
Models	GloVe + CNN F1	GloVe + CNN AUC	BERT F1	BERT AUC
Proto-HATT♣	57.26	90.63	57.33	89.70
LDF-HATT	60.68±0.92	91.22±0.53	63.72±0.27	91.99±0.12
Proto-AWATT♣	65.65	92.06	70.09	94.59
LDF-AWATT	67.13±0.41	92.74±0.12	72.76±0.29	95.31±0.19

Table 5: The effect of different encoders in the 10-way 5-shot scenario on FewAsp dataset. The results with symbol ♣ are retrieved from (Hu et al., 2021).

Models	10-way 5-shot F1	10-way 5-shot AUC
Proto-AWATT	65.65	92.06
Proto-AWATT (LSW)	57.84±0.49	90.85±0.22

Table 6: The effect of label similarity weight α in the 10-way 5-shot scenario on FewAsp dataset.

Experiments



(a) Proto-AWATT (b) Proto-AWATT+LAS (c) LDF-AWATT

Figure 5: Visualization of prototype representations for Proto-AWATT, Proto-AWATT+LAS and LDF-AWATT.

Thank you!