



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

Fei zhao
zhaof@smail.nju.edu.cn



提纲



- 任务介绍
- 动机与方法
- 实验结果
- 总结贡献



任务介绍



- 多标签少样本方面类别检测（FS-ACD）：旨在有限的监督数据下检测方面类别

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



现有方法

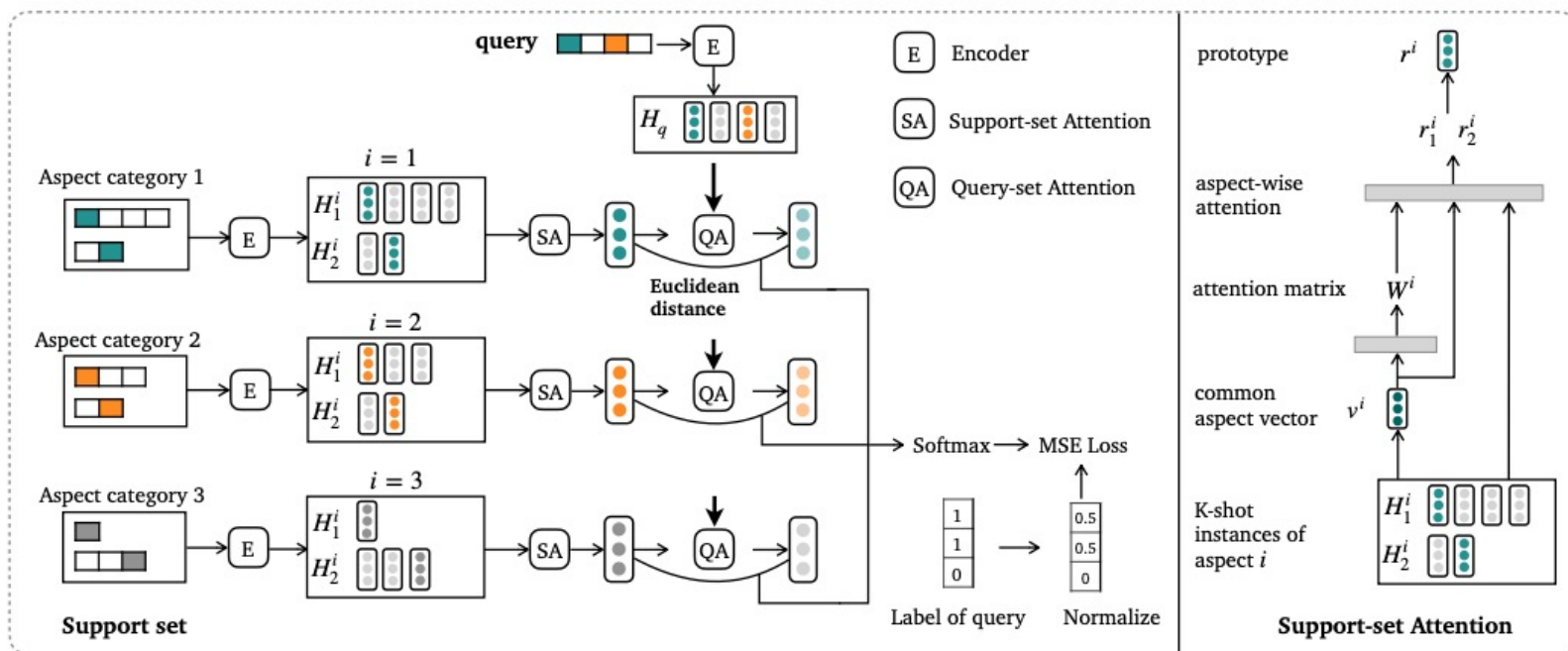


图1: Proto-AWATT框架



动机：噪音问题



- 一方面：由于缺乏足够的监督数据，以前的模型很容易捕捉到与当前方面类别无关的噪音词，这在很大程度上影响了生成原型的质量

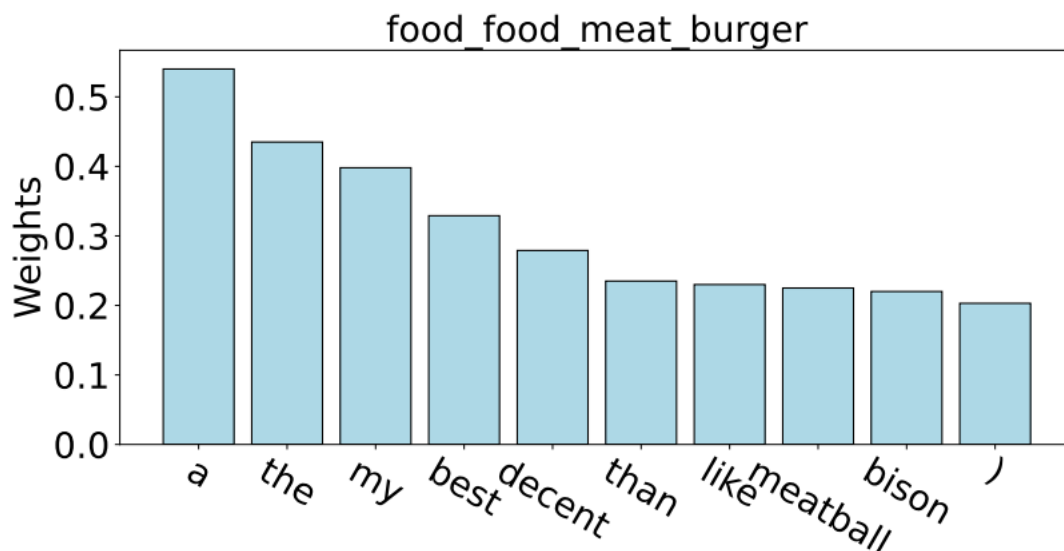


图2: 可视化方面类别 food_food_meat_burger 原型的前 10 个单词



动机：噪音问题



- 另一方面：语义上接近的方面类别通常会产生相似的原型，这些相似的原型互为噪音，极大地混淆了分类器

| 方面类别对的关系 | Percentage (%) |
|---------------|----------------|
| 方面类别对具有相似的语义 | 25.50% |
| 方面类别对不具有相似的语义 | 74.50% |



标签驱动去噪框架

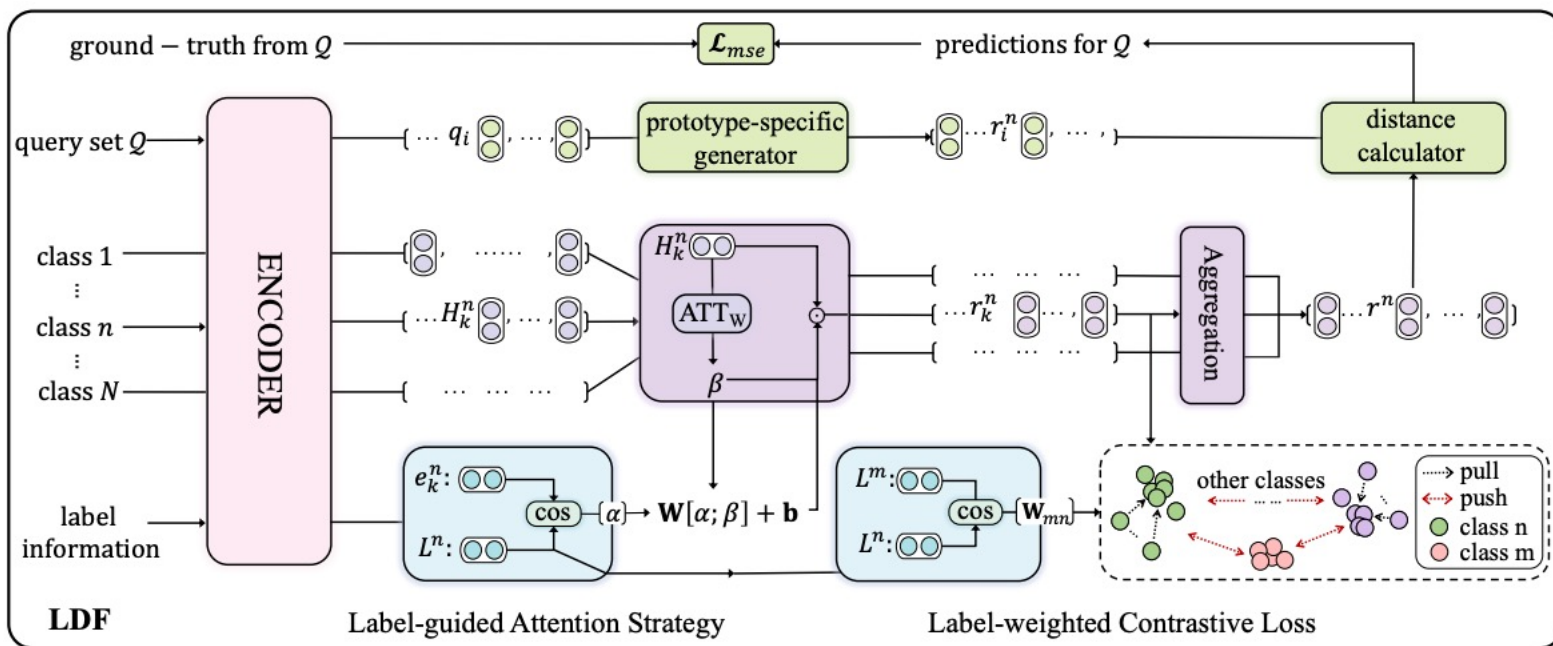


图3: 标签驱动去噪框架LDF的整体架构



实验设置



- 基线方法
 - Proto-HATT 和 Proto-AWATT
- 数据集
 - FewAsp(single)、FewAsp(multi) 和 FewAsp
- 评测指标
 - Macro-F1 和 AUC
 - 5 way 和 10 way的阈值分别设置为 $\{0.3, 0.2\}$
 - 我们报告五次运行后的平均性能和标准差，其中种子设置为 $[5, 10, 15, 20, 25]$



实验结果



主实验结果

| Models | 5-way 5-shot | | 5-way 10-shot | | 10-way 5-shot | | 10-way 10-shot | |
|-----------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| | F1 | AUC | F1 | AUC | F1 | AUC | F1 | AUC |
| <i>FewAsp</i> | | | | | | | | |
| 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 |
| <i>FewAsp(single)</i> | | | | | | | | |
| 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 |
| <i>FewAsp(multi)</i> | | | | | | | | |
| 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 |



实验结果



消融实验结果

| Models | 5-way 5-shot | | 5-way 10-shot | | 10-way 5-shot | | 10-way 10-shot | |
|------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| | F1 | AUC | F1 | AUC | F1 | AUC | F1 | 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 |



■ 编码器的影响

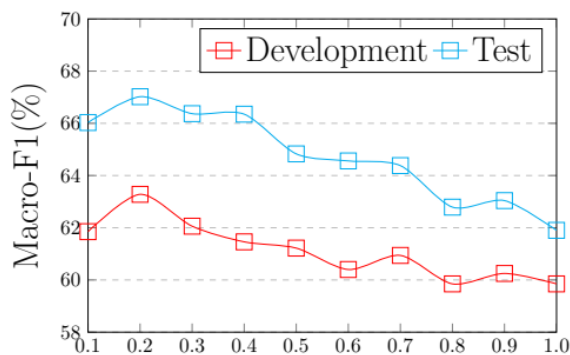
| Models | GloVe + CNN | | BERT | |
|--------------------------|-------------|------------|------------|------------|
| | F1 | AUC | F1 | AUC |
| Proto-HATT [♣] | 57.26 | 90.63 | 57.33 | 89.70 |
| LDF-HATT | 60.68±0.92 | 91.22±0.53 | 63.72±0.27 | 91.99±0.12 |
| Proto-AWATT [♣] | 65.65 | 92.06 | 70.09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

■ 标签相似度权重的影响

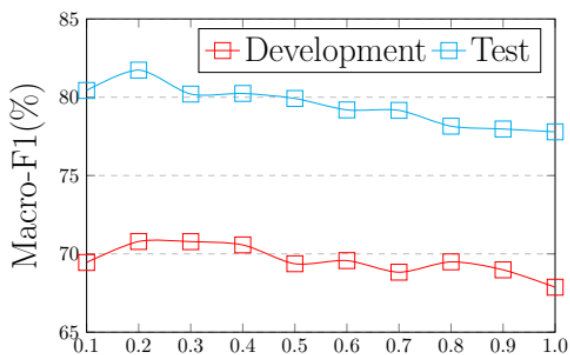
| Models | 10-way 5-shot | |
|-------------------|---------------|------------|
| | F1 | AUC |
| Proto-AWATT | 65.65 | 92.06 |
| Proto-AWATT (LSW) | 57.84±0.49 | 90.85±0.22 |



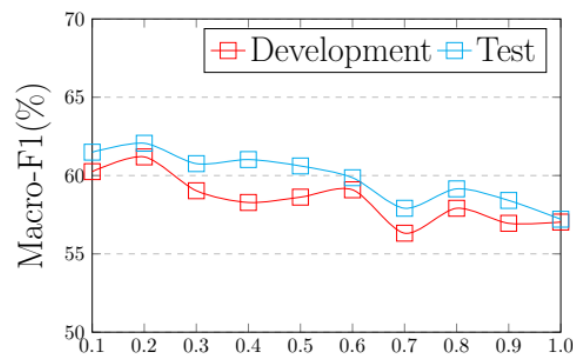
超参数的影响



(a) FewAsp



(b) FewAsp(single)

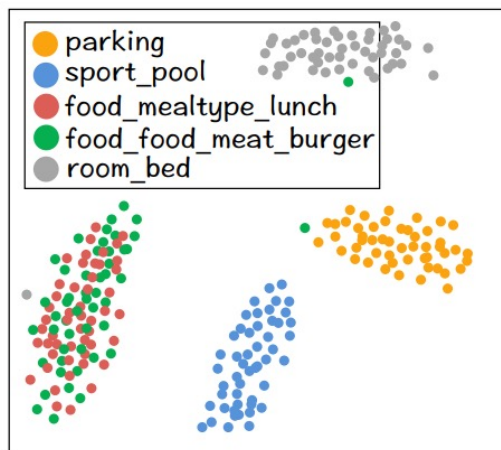


(c) FewAsp(multi)

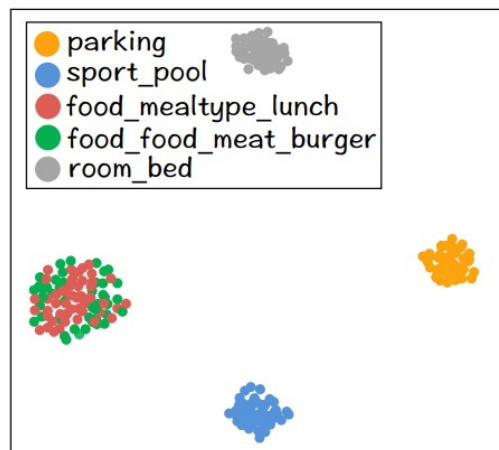
图4: 超参数在 10 way 5-shot 设置中对三个数据集的影响



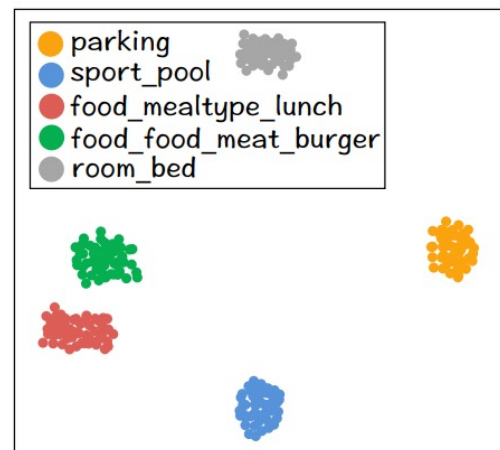
■ 案例分析



(a) Proto-AWATT



(b) Proto-AWATT+LAS



(c) LDF-AWATT

图5: Proto-AWATT、Proto-AWATT+LAS 和 LDF-AWATT 原型表示的可视化



分析与讨论



■ 错误分析

| Category | Proportion | Example | True Label | Predict Label |
|------------------|------------|---|---|-----------------------|
| Complex | 41% | (1) fast forward to december 2014, we have a company gathering in one of the many banquet rooms at the chandler downtown serrano. | restaurant_location | room_interior ✗ |
| No obvious clues | 22% | (2) overall, this is a great salon, and I will be back ! | procedure_beauty_nails experience_wait | salon_interior_room ✗ |



总结贡献



- 据我们所知，我们是第一个利用每个类的标签信息来提高FS-ACD任务性能的工作
- 我们引入了一种标签引导的注意力策略为每个方面类别生成具有代表性的原型，以及一种标签加权对比学习方法来区分语义上接近的原型
- 广泛的实验结果表明，我们的框架在三个public数据集上始终优于当前最先进的工作



谢谢！
请各位老师同学批评指正！