

Metric-Free Learning Network with Dual Relations Propagation for Few-Shot Aspect Category Sentiment Analysis

Shiman Zhao, Yutao Xie, Wei Chen*, Tengjiao Wang,
Jiahui Yao, Jiabin Zheng

Key Lab of High Confidence Software Technologies (MOE),
School of Computer Science, Peking University, Beijing, China
Research Center for Computational Social Science, Peking University
Institute of Computational Social Science, Peking University (Qingdao)
{shimanzhao, yutao_xie}@stu.pku.edu.cn
{pekingchenwei, tjwang, issyaojh, jiabinzheng}@pku.edu.cn

Abstract

Few-shot Aspect Category Sentiment Analysis (ACSA) is a crucial task for aspect-based sentiment analysis, which aims to detect sentiment polarity for a given aspect category in a sentence with limited data. However, few-shot learning methods focus on distance metrics between the query and support sets to classify queries, heavily relying on aspect distributions in the embedding space. Thus, they suffer from overlapping distributions of aspect embeddings caused by irrelevant sentiment noise among sentences with multiple sentiment aspects, leading to misclassifications. To solve the above issues, we propose a metric-free method for few-shot ACSA, which models the associated relations among the aspects of support and query sentences by Dual Relations Propagation (DRP), addressing the passive effect of overlapping distributions. Specifically, DRP uses the dual relations (similarity and diversity) among the aspects of support and query sentences to explore intra-cluster commonality and inter-cluster uniqueness for alleviating sentiment noise and enhancing aspect features. Additionally, the dual relations are transformed from support-query to class-query to promote query inference by learning class knowledge. Experiments show that we achieve convincing performance on few-shot ACSA, especially an average improvement of 2.93% accuracy and 2.10% F1 score in the 3-way 1-shot setting.

1 Introduction

Aspect Category Sentiment Analysis (ACSA) (Seoh et al., 2021; Cai et al., 2021; Xiao et al., 2021; Chen et al., 2022a; Li et al., 2022a,b) is a

fine-grained sentiment analysis task, which aims to identify sentiment polarity for a given aspect category in a sentence. For example, given a predefined aspect category “*Staff*” and a sentence “*High rates for just ok room but the server keeps me waiting 1.5 hours*”, ACSA aims to identify sentiment polarity towards the aspect “*Staff*” in the sentence. Briefly, given an aspect category and a sentence, an aspect embedding is obtained from the original sentence to predict the sentiment polarity of the aspect category in the sentence.

Existing methods mostly rely on sufficient labeled data for each aspect category. Though effective, they assume training and testing share a predefined set of aspects. However, this assumption becomes problematic in real-world scenarios with many unseen aspect categories. Annotating abundant data for these emerging aspects poses a significant challenge, and there is a burden of retraining models for newly encountered aspects. Therefore, generalizing experiences from seen aspect categories to unseen ones becomes crucial. This is where few-shot ACSA becomes indispensable.

Existing few-shot learning methods (e.g., meta-learning) mostly focus on distance metrics (Yang et al., 2020; Wang et al., 2021; Lv et al., 2021; Assran et al., 2022; Liu et al., 2022a). Among these methods, the prototypical network is a distance metric method well known because of its impressive performance. The prototypical network uses the support set to generate a prototype for each class and then classifies the query by measuring the distance (e.g., Euclidean distance or cosine similarity) with different prototypes in the embedding space.

*Corresponding author.

Support Set			
Task ID	Aspect	Sentence	Sentiment
1	Staff	Although he looks fierce, the clerk is actually friendly.	Positive
		The staff keeps me waiting 2 hours but the decor looks pretty.	Negative
2	Price	Promotional activities could boost sales growth.	Positive
		With low supply, the battery is more expensive.	Negative
Query Set			
Aspect	Sentence		Sentiment
Staff	The buffet tastes great although the service is slack.		Negative

Table 1: An episode consisting of two meta-tasks in a 2-way 1-shot setting during the meta-test phase. The color highlights the sentiment contexts for aspects “Staff” and “Price”. Grey words are irrelevant contexts, which can be considered as noise. In the query set, the dashed box indicates the input of the query (i.e., aspect and sentence), whereas the full box denotes its expected output (i.e., sentiment label).

Though few-shot learning achieves impressive progress, there are challenging issues for the few-shot ACSA task. Specifically, **simple distance metrics struggle to address overlapping distributions of aspect embeddings** caused by irrelevant sentiment noise in scenarios (e.g., Table 1) where each sentence may include many aspects with different sentiment polarities. Generally, overlapping distributions present an unclear decision boundary among aspect embeddings with different sentiment polarities, causing misclassifications. Take the example in Figure 1; the aspects “Service” and “Price” show a closer distance than “Service” and other aspects, indicating they tend to have the same sentiment polarity. However, in reality, they should have opposite sentiment polarities, resulting in final wrong predictions. Recent efforts have been devoted to these issues. Liang et al. (2023) explored external knowledge (e.g., aspect-associated words and aspect semantics) to alleviate irrelevant sentiment noise for enhancing aspect embeddings. However, maintaining and updating the knowledge base requires domain experts, making it resource-intensive. Additionally, collecting abundant knowledge for unseen aspect categories limits scalability. Therefore, the mentioned issues are still a considerable challenge. As the research on few-shot ACSA is still young, a novel method is expected to perform on the few-shot ACSA task.

To solve the above issues, we propose a metric-free method to address the few-shot ACSA

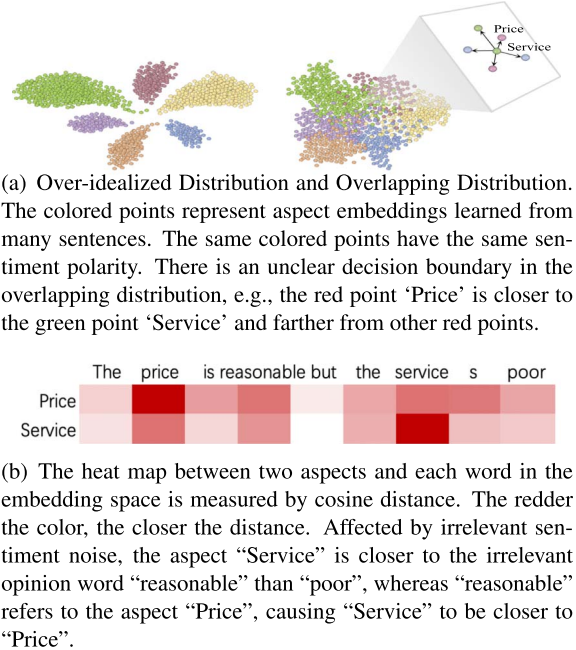


Figure 1: Subfigure(a) shows an over-idealized distribution and an overlapping distribution. Subfigure(b) gives the reason for the overlapping distribution.

task by modeling Dual Relations Propagation (DRP). Following the meta-learning formulation, DRP is designed in a relation graph to explicitly model the dual relations (i.e., similarity relation¹ and diversity relation²) among support and query

¹Similarity relation indicates that nodes from the same sentiment label share similar sentiment features.

²Diversity relation indicates that nodes from different sentiment labels express contrasting sentiment features.

nodes. In the relation graph, each node is a sentence-aspect pair, and its aspect embedding is considered as the node feature in the embedding space. Additionally, the dual relations are formalized as two undirected edges between a node pair, and each relation has an associated strength measure. Briefly, the similarity relation presents a similarity strength between a connected node pair, and the diversity relation gives a discrepancy strength between them. With the relation graph, the proposed method propagates and aggregates the dual relations to explore intra-cluster commonality and inter-cluster uniqueness, alleviating irrelevant sentiment noise and enhancing node and edge features. Also, the dual relations are transformed from support-query to class-query to guide query inference by learning sentiment class knowledge from the relation graph. Extensive experiments show that the proposed method outperforms strong baselines and obtains significant performance. Significantly, it surpasses the latest baseline by 2.93% accuracy and 2.10% F1 score on average in the 3-way 1-shot setting. The contributions are summarized as follows:

- An effective metric-free method for the few-shot ACSA task is proposed by modeling dual relations propagation. The dual relations propagation exploits the similarity and diversity among the support and query sets to explore intra-cluster commonality and inter-cluster uniqueness to address the passive effect of overlapping distributions caused by irrelevant sentiment noise.
- The proposed method transforms the dual relations from support-query to class-query to promote query inference by learning sentiment class knowledge.
- Extensive experiments on four benchmark datasets show that the proposed method outperforms strong baselines and obtains significant performance on few-shot ACSA.

2 Related Work

2.1 Aspect Category Sentiment Analysis

The ACSA task aims to detect sentiment polarity for a specific aspect mentioned in a sentence. Generally, it is used in recommendation systems (Cui et al., 2020; Jannach et al., 2021; Ahmadian

et al., 2022) and intention detection (Hou et al., 2021; Chen et al., 2022c; Zhou et al., 2022) to understand the fine-grained sentiment of users. In recent years, ACSA has attracted the attention of researchers and developers.

Conventional methods focus on handcraft-based and attention-based methods. Handcraft-based methods (Ding et al., 2015; Liu et al., 2015) utilize handcrafted features to establish the dependency between a specific aspect and its context. Attention-based methods (Su et al., 2021; Wu et al., 2021; Liu et al., 2021) capture the interaction between an aspect and its context. Recently, some syntax-aware methods (Tian et al., 2021; Li et al., 2021b; Xiao et al., 2022; Effland and Collins, 2023) utilized Graph Neural Networks (GNN) based on syntactical dependency trees to exploit syntactic structure information. However, these methods heavily rely on labeled data and may fail to solve unseen aspect categories. Therefore, few-shot learning is of great importance.

2.2 Few-Shot Learning

Few-shot learning (Tsendsuren and Hong, 2017; Lee et al., 2019b; Zhang et al., 2022a) matches the human learning process in that the few-shot learner leverages a few labeled samples to obtain new knowledge based on prior knowledge. Few-shot learning has achieved promising processes in Computer Vision (CV) (Huang et al., 2021; Hu et al., 2022; Liu et al., 2022b; Ouyang et al., 2022), Natural Language Processing (NLP) (Hu et al., 2021; Tan et al., 2022; Chen et al., 2022b; Gao et al., 2022), etc. Especially in NLP, a number of research works exist on few-shot learning, such as few-shot aspect category detection (Zhao et al., 2023), few-shot named entity recognition (Fang et al., 2023; Xu et al., 2023; Ma et al., 2023), few-shot relation extraction (Chen et al., 2023; Li et al., 2023), etc.

Few-shot learning mainly contains meta-learning, prompt learning, and data augmentation. Meta-learning (Sung et al., 2018) leverages prior experiences to enable the model to obtain learning abilities and generalize them to new fields. Prompt learning (Lu et al., 2022) constructs task-related prompts to guide large language models to generate task-specific outputs. Data augmentation (Zhang et al., 2020) transforms existing samples to expand the dataset to promote the model to learn data patterns and features.

2.3 Meta-Learning

In recent years, meta-learning has been the main few-shot learning method due to its impressive performance, including model-based (Tsendsuren and Hong, 2017), optimization-based (Lee et al., 2019b), and metric-based (Assran et al., 2022; Wang et al., 2021; Lv et al., 2021) methods. Among them, metric-based methods are the most popular research line on meta-learning due to their simplicity and effectiveness. The main idea (Yu et al., 2022; Zhang et al., 2022b) is to use an episode paradigm to project support and query samples to an embedding space and then measure their distances to predict query labels. However, these methods heavily rely on aspect distributions in the embedding space. Therefore, they suffer from overlapping distributions of aspect embeddings caused by irrelevant sentiment noise among sentences with different sentiment aspects. Recently, Hosseini-Asl et al. (2022) proposed a generative method to explore aspect semantics to capture the interactions between a specific aspect and its context. More recently, Liang et al. (2023) leveraged aspect-associated words from an external knowledge base (Cambria et al., 2020) to construct two auxiliary sentences to enhance aspect embeddings. Unfortunately, their improvements are limited due to the complexity of semantic relations and knowledge structures. Therefore, these methods still struggle to handle irrelevant sentiment noise in scenarios where each sentence contains different sentiment aspects.

Unlike the mentioned methods, the proposed method explores the shared features among samples in a class and diverse features in separate classes to model the dual relations (similarity and diversity) among samples. With relation propagation and aggregation, the proposed method alleviates irrelevant sentiment noise and enhances sample features to improve performance on the few-shot ACSA task. Compared to previous methods, the proposed method works well in scenarios where each sentence contains different sentiment aspects.

3 Proposed Method

The overall architecture of the proposed method is shown in Figure 2. Broadly, the proposed method includes four components: relation graph construction, support-query relation propagation, class-query relation transformation, and training

objective. Here, we present the proposed method in detail.

3.1 Problem Formulation

Following the meta-learning formulation, we handle the few-shot ACSA task in an episode paradigm. Meta-learning takes the meta-task as the basic unit and constructs episodes to train/test the model in meta-train, meta-val, and meta-test phases. An episode consists of meta-tasks, and each meta-task consists of a support set \mathcal{S} and a query set \mathcal{Q} for a specific aspect. Therefore, we divide the aspects into \mathcal{T}_{train} , \mathcal{T}_{val} , and \mathcal{T}_{test} , where $\mathcal{T}_{train} \cap \mathcal{T}_{val} \cap \mathcal{T}_{test} = \emptyset$. The segmentation strategy generalizes sentiment knowledge from seen aspects to emerging unseen aspects. Formally, we suppose that Λ^a aspects mentioned in the dataset are written as:

$$\mathcal{T}_{train} = \{\tau_1, \tau_2, \dots, \tau_{\Lambda^t}\}, \quad (1)$$

$$\mathcal{T}_{val} = \{\tau_{\Lambda^t+1}, \tau_{\Lambda^t+2}, \dots, \tau_{\Lambda^g}\}, \quad (2)$$

$$\mathcal{T}_{test} = \{\tau_{\Lambda^g+1}, \tau_{\Lambda^g+2}, \dots, \tau_{\Lambda^a}\}. \quad (3)$$

In the meta-train phase, the model is trained on \mathcal{T}_{train} through episodes. An episode consists of meta-tasks, and each meta-task is related to an aspect. Specifically, the meta-learner extracts a meta-task for each aspect in \mathcal{T}_{train} to construct the episode. For a meta-task, the meta-learner first utilizes a N -way K -shot setting to construct \mathcal{S} , i.e., there are N classes (i.e., N sentiment labels), and each class has K samples. Then, T samples are randomly selected from the remaining samples of the N classes to construct the query set \mathcal{Q} . Meta-task aims to use \mathcal{S} to classify query samples and minimize the loss of query prediction. We use x , a , and y to define the support set and the query set, where x and a represent the sentence and its aspect, and y is the sentiment label of the aspect a ($a \in \mathcal{T}_{train}$) in the sentence x . In an episode, \mathcal{S} and \mathcal{Q} could be written as:

$$\mathcal{S} := \{(x_i^s, a_i^s, y_i^s)\}, i \in [1, |\mathcal{S}|], \quad (4)$$

$$\mathcal{Q} := \{(x_i^q, a_i^q)\}, i \in [1, |\mathcal{Q}|], \quad (5)$$

where $|\mathcal{S}| = N \times K \times \Lambda^t$, and $|\mathcal{Q}| = N \times T \times \Lambda^t$. N is the number of labels, K and T denote the number of samples for each label. Also, Λ^t is associated with the number of meta-tasks and could be realized as the number of aspects in \mathcal{T}_{train} , as shown in Equation 1.

In the meta-val or meta-test phase, the meta-learner aims to verify the effectiveness of the

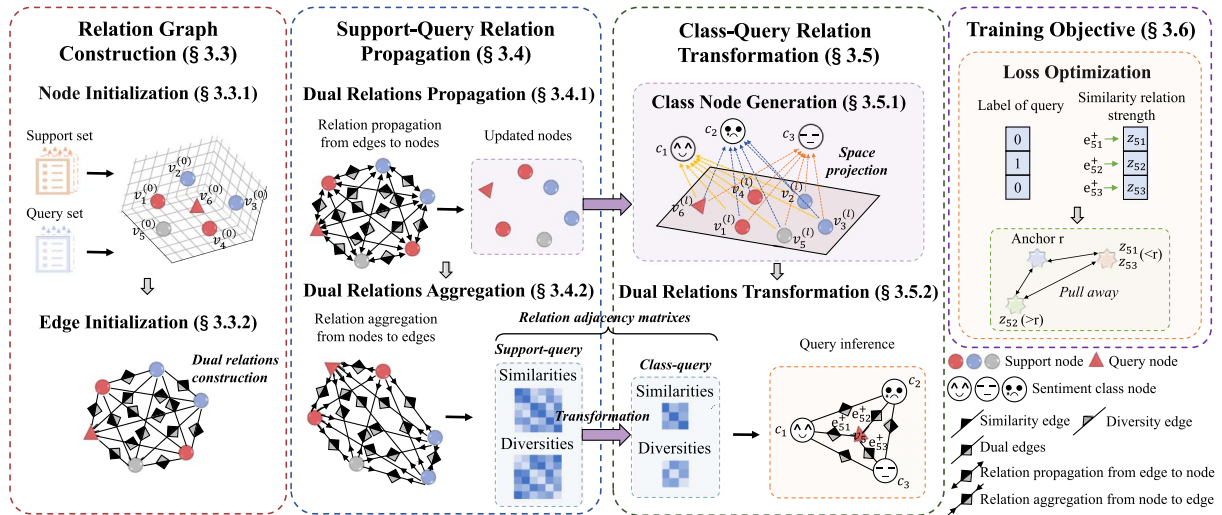


Figure 2: The overall architecture of the proposed method. The relation graph consists of support (circle) and query (triangle) nodes, and different colors mark different sentiment polarities of these nodes. In the relation graph, the arrow direction of edges represents the relation propagation or aggregation. In class-query relation transformation, emoji nodes (c_1 , c_2 , and c_3) are considered as sentiment class nodes.

model in $\mathcal{T}_{val}/\mathcal{T}_{test}$. Unlike the meta-train phase, the meta-learner only constructs a fixed support set \mathcal{S} for each aspect based on $\mathcal{T}_{val}/\mathcal{T}_{test}$ and then takes the remaining samples in the dataset as the query set, as shown in Table 1. The meta-learner aims to use \mathcal{S} to predict the labels of query samples in \mathcal{Q} and evaluate the performance of the proposed method. Finally, we report the corresponding results of the meta-test phase when the meta-val phase obtains the best results.

3.2 Overall Framework

As shown in Figure 2, the proposed method includes four components: relation graph construction, support-query relation propagation, class-query relation transformation, and training objective. Specifically, the proposed method designs a simple yet effective relation graph. The relation graph is an undirected, fully connected graph that aims to model dual relations (i.e., similarity and diversity) among support and query nodes. In the relation graph, each node is a sentence-aspect pair, and its aspect embedding (see Equation 7) is considered as the node feature. Additionally, two edges are used to connect two nodes, and edge features indicate the similarity and diversity strength between these two nodes.

With the relation graph, the proposed method propagates dual relations to enrich node features from edges to nodes and aggregates dual rela-

tions to update edge features from nodes to edges. Briefly, the propagation and aggregation of the dual relations enhance node and edge features and alleviate irrelevant sentiment noise by exploring intra-cluster commonality and inter-cluster uniqueness. Besides, the dual relations are transformed from support-query to class-query to promote query inference effectively.

3.3 Relation Graph Construction

We present the relation graph construction, including node initialization and edge initialization, as shown on the left of Figure 2. The relation graph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E}^+, \mathcal{E}^-)$, where $\mathcal{V} = \{V_i\}_{i=1}^M$ denotes the set of nodes, $\mathcal{E}^+ = \{E_{ij}^+\}_{i,j}^M$ and $\mathcal{E}^- = \{E_{ij}^-\}_{i,j}^M$ indicate similarity edges and diversity edges, respectively, and M is the total number of nodes. Briefly, two edges as dual bridges are built between two adjacent nodes to represent the similarity and diversity relations of the node pair. Besides, v_i represents the features of node V_i , and e_{ij}^+ and e_{ij}^- indicate the features of edges E_{ij}^+ and E_{ij}^- . The following initialization of nodes and edges is presented in the relation graph.

3.3.1 Node Initialization

Given a sentence-aspect pair (x, a) , the sentence x with n words is defined as $x = \{w_1, w_2, \dots,$

$w_n\}$, and the aspect $a = \{s_1, s_2, \dots, s_m\}$ consists of m words. The sentence and aspect are concatenated to construct “[CLS], x , [SEP], a , [SEP]” as an input to an encoder (e.g., BERT [Devlin et al., 2019]) to generate hidden states H .

$$H = \text{Encoder}([CLS, x, SEP, a, SEP]), \quad (6)$$

where $H \in R^{(n+m+3) \times d}$ denotes the hidden states of the input, and d is the dimension of hidden states. Then, a mean pooling layer obtains the aspect embedding as node features.

$$v^{(0)} = \text{MeanPoolLayer}(H), \quad (7)$$

where $v^{(0)} \in R^d$ indicates a node feature projected to an embedding space in the 0^{th} layer.

3.3.2 Edge Initialization

Between a node pair V_i and V_j , two edges E_{ij}^+ and E_{ij}^- represent the similarity and diversity relations of the two nodes, and the edge features e_{ij}^+ and e_{ij}^- represent the strength of similarity and diversity relations. Briefly, e_{ij}^+ is a probability that nodes V_i and V_j from the same class, while e_{ij}^- is the probability that they belong to different classes.

Therefore, node labels could be used to initialize dual-edge features by exploiting intra-cluster commonality and inter-cluster uniqueness.

$$e_{ij}^{+(0)}/e_{ij}^{-+(0)} = \begin{cases} 1/-1, & \text{if } y_i = y_j, \{V_i^{(0)}, V_j^{(0)}\} \in \mathcal{S} \\ -1/1, & \text{if } y_i \neq y_j, \{V_i^{(0)}, V_j^{(0)}\} \in \mathcal{S} \\ 0/0, & \text{otherwise} \end{cases} \quad (8)$$

where y_i is the label of node $V_i^{(0)}$.

In the relation graph, for inter-cluster nodes, the lower the similarity strength, the greater the difference between them; conversely, for intra-cluster nodes, the lower the diversity strength, the more common features they share.

3.4 Support-Query Relation Propagation

The relation graph promotes the modeling of support-query relations. As shown in Figure 2, the component includes dual relations propagation and aggregation, which aims to learn discriminative node and edge features.

3.4.1 Dual Relations Propagation

The dual relations propagation enriches node features from dual edges to nodes by learning neighborhood knowledge. Specifically,

$$\tilde{v}_i^{(l)} = \left[\sum_j \frac{e_{ij}^{+(l-1)}}{\sum_k e_{ik}^{+(l-1)}} v_j^{(l-1)}; \sum_j \frac{e_{ij}^{-(l-1)}}{\sum_k e_{ik}^{-(l-1)}} v_j^{(l-1)} \right], \quad (9)$$

$$v_i^{(l)} = W_1 \text{ReLU}(\text{MLP}(\tilde{v}_i^{(l)}) + v_i^{(l-1)}) + b_1, \quad (10)$$

where $v_i^{(l)}$ represents the i^{th} node feature in the l^{th} layer. $\tilde{v}_i^{(l)}$ utilizes normalized dual-edge features to merge similarity and diversity node features. $\text{MLP}(\ast)$ is a multi-layer perception, and $\text{ReLU}(\ast)$ is an activation function. $[\ast; \ast]$ denotes the concatenation between two vectors. W_1 and b_1 are trainable parameters.

3.4.2 Dual Relations Aggregation

The dual relations aggregation uses the latest node features to update edge features to learn robust dual relations. Specifically,

$$B = \text{LayerNorm}(\text{MLP}(v_i^{(l)} - v_j^{(l)})), \quad (11)$$

$$\tilde{e}_{ij}^{+(l)} = \text{MLP}(\text{ReLU}(B + \max(v_i^{(l)}, v_j^{(l)}))), \quad (12)$$

$$\tilde{e}_{ij}^{-(l)} = \text{MLP}(\text{ReLU}(B + \min(v_i^{(l)}, v_j^{(l)}))), \quad (13)$$

where $\text{LayerNorm}(\ast)$ is a layer normalization, and $\max(\ast, \ast)$ and $\min(\ast, \ast)$ are element-wise operations. $\tilde{e}_{ij}^{+(l)}$ and $\tilde{e}_{ij}^{-(l)}$ denote the similarity and diversity edge features between nodes $V_i^{(l)}$ and $V_j^{(l)}$ in the l^{th} layer. Then, dual-edge features are normalized by other edges connected to the node.

$$\hat{e}_{ij}^{+(l)} = \frac{\tilde{e}_{ij}^{+(l)}}{\sum_k \tilde{e}_{ik}^{+(l)}}, \quad \hat{e}_{ij}^{-(l)} = \frac{\tilde{e}_{ij}^{-(l)}}{\sum_k \tilde{e}_{ik}^{-(l)}}, \quad (14)$$

$$e_{ij}^{+(l)} = \frac{\hat{e}_{ij}^{+(l)}}{\|[\hat{e}_{ij}^{+(l)}; \hat{e}_{ij}^{-(l)}]\|_2}, \quad e_{ij}^{-(l)} = \frac{\hat{e}_{ij}^{-(l)}}{\|[\hat{e}_{ij}^{+(l)}; \hat{e}_{ij}^{-(l)}]\|_2}, \quad (15)$$

where $\| \ast \|_2$ denotes L2 norm.

During relation propagation, the sentiment label of a query node can be predicted by final edge voting with support labels. However, edge voting makes the prediction difficult because the relation graph contains many edges. Therefore, the proposed method transforms support-query relations into class-query relations to promote query inference effectively.

3.5 Class-Query Relation Transformation

As shown in Figure 2, the proposed method transforms dual relations from support-query to class-query by learning sentiment class knowledge, which models the relations between a query and different sentiment classes to promote query inference further.

3.5.1 Class Node Generation

First, the proposed method generates class node features from the features of support and query nodes by space projection.

$$P = \text{softmax}(W_2 V^{(l)} + b_2), \quad (16)$$

$$V_c^{(l)} = P^T V^{(l)}, \quad (17)$$

where $V^{(l)} \in R^{M \times d}$ denotes the original node set in the l^{th} layer, and $V_c^{(l)} \in R^{G \times d}$ indicates the sentiment class node set. M and G are the number of original nodes and sentiment class nodes, respectively. In Figure 2, we use c_i to represent each class node feature vector. Additionally, P is a probability matrix with size $M \times G$ from original nodes to sentiment class nodes. W_2 and b_2 are trainable parameters. The $\text{softmax}(\ast)$ function ensures that elements in each row of P are in the range $[0, 1]$, and the sum of elements in each row is 1.

3.5.2 Dual Relations Transformation

The dual relations transformation from support-query to class-query is written as:

$$E_c^{+/- (l)} = P^T E^{+/- (l)} P, \quad (18)$$

where $E^{+/- (l)}$ is the similarity/diversity relation adjacency matrix with size $M \times M$ for support and query nodes. $E_c^{+/- (l)}$ is the similarity/diversity relation adjacency matrix with size $G \times G$ for class and query nodes. Therefore, the label of

a query can be predicted by the sentiment label of the class with the strongest similarity to the query.

3.6 Training Objective

During the meta-train phase, we suppose there are k sentiment classes. Given a query q_i , the strength score of the similarity relation between it and sentiment classes is defined as $Z_i = \{z_{i1}, z_{i2}, \dots, z_{ik}\} \in R^k$, i.e., $z_{ij} = e^{z_{ij}^{(l)}}$ ($j \in \{1, 2, \dots, k\}$). The one-hot label of q_i is defined as $y_i = \{y_{i1}, y_{i2}, \dots, y_{ik}\} \in \{0, 1\}^k$, where $y_{ij} = 1$ indicates q_i belongs to j^{th} class. For training, we define the positive set $\Omega_i^{pos} = \{z_{ij} \in Z_i | y_{ij} = 1\}$ and the negative set $\Omega_i^{neg} = \{z_{ij} \in Z_i | y_{ij} = 0\}$.

For the query q_i , the contrastive training objective (Su et al., 2022) minimizes the following loss function.

$$\begin{aligned} \mathcal{L}_i &= \log(1 + \sum_{\substack{z_{ik} \in \Omega_i^{neg} \\ z_{ij} \in \Omega_i^{pos}}} e^{z_{ik} - z_{ij}} + \sum_{z_{ik} \in \Omega_i^{neg}} e^{z_{ik} - r} + \sum_{z_{ij} \in \Omega_i^{pos}} e^{r - z_{ij}}) \\ &= \log(e^r e^{-r} + \sum_{z_{ik} \in \Omega_i^{neg}} e^{z_{ik}} \sum_{z_{ij} \in \Omega_i^{pos}} e^{-z_{ij}} + e^{-r} \sum_{z_{ik} \in \Omega_i^{neg}} e^{z_{ik}} \\ &\quad + e^r \sum_{z_{ij} \in \Omega_i^{pos}} e^{-z_{ij}}) \\ &= \log((e^r + \sum_{z_{ik} \in \Omega_i^{neg}} e^{z_{ik}})(e^{-r} + \sum_{z_{ij} \in \Omega_i^{pos}} e^{-z_{ij}})) \\ &= \log(e^r + \sum_{z_{ik} \in \Omega_i^{neg}} e^{z_{ik}}) + \log(e^{-r} + \sum_{z_{ij} \in \Omega_i^{pos}} e^{-z_{ij}}), \end{aligned} \quad (19)$$

where r is an anchor. The optimization objective \mathcal{L}_i is to ensure that the scores in Ω_i^{pos} are greater than r and the scores in Ω_i^{neg} are less than r . We set $r = 0$ to promote the strength of the similarity relation in intra-cluster are positive and in inter-cluster are negative.

Given a query set \mathcal{Q} , the overall training objective is as follows.

$$\mathcal{L} = \frac{1}{|\mathcal{Q}|} \sum_{i=1}^{|\mathcal{Q}|} \mathcal{L}_i, \quad (20)$$

where $|\mathcal{Q}|$ is the number of queries.

During the meta-test phase, we use the similarity strength score Z_i ($Z_i = \{z_{i1}, z_{i2}, \dots, z_{ik}\} \in R^k$) between the query q_i and different sentiment classes to predict the label \hat{y}_i of q_i . The higher the similarity score, the closer the distance.

$$\hat{y}_i = \underset{j}{\text{argmax}} z_{ij}, \{z_{ij} > 0 | z_{ij} \in Z_i\}. \quad (21)$$

Datasets	Rest I	Res II	Lap	Mams
#Asp.	4	8	12	8
#Pos.	2596	1825	1499	2415
#Neu.	583	105	112	3863
#Neg.	1145	823	950	2606

Table 2: Statistics of four datasets. #Asp. denotes the number of aspect categories in datasets. #Pos., #Neu., and #Neg. indicate the number of positive, neutral, and negative sentiment polarities.

4 Experiments

4.1 Experimental Setup

Datasets. Extensive experiments are conducted on four datasets: Rest I, Rest II, Lap, and Mams. These four datasets were collected by Liang et al. (2023) for the few-shot ACSA task. Rest I and Rest II originate from the restaurant domain, with Rest II providing a fine-grained aspect (entity #attribute) compared to Rest I (entity). Lap is obtained from the laptop domain to explore performance in other domains. For these three datasets, most sentences contain only one or multiple aspects with the same sentiment polarity. Therefore, Mams presents a more complex scenario, where each sentence includes many aspects with different sentiment polarities. The detailed statistics are presented in Table 2. Our code and data are available at <https://github.com/sentiments-Ananda/FSACSA>.

Evaluation Metric. Following previous methods (Li et al., 2021a; Liang et al., 2023), we use accuracy and F1 score to evaluate and compare the performance of the proposed method.

Implementation Details. The proposed method is implemented with PyTorch (version 1.10.0). The uncased English version of BERT is our encoder for H (see Equation 6). In practice, the bottom layers of large language models are unnecessary (Lee et al., 2019a). Thus, we freeze the first six layers of BERT to reduce trainable parameters. We conduct experiments on a single GPU (RTX 3090 Ti) with CUDA version 11.3. The model is trained by the AdamW optimizer. To ensure a fair comparison, we follow Liang et al. (2023) to obtain experiment results by using a four-fold cross-validation. For example, a

dataset has eight aspects, and these aspects are divided into four folds. We take each fold as the testing set and the others as the validation and training sets, and the splitting proportion is 1:1:2 for testing, validation, and training. The schematic diagram of the four-fold cross-validation is shown in Figure 3. Therefore, we can obtain four experiment results, and the average of these four experiment results is calculated to evaluate the performance of the proposed method.

4.2 Baselines

We compare the proposed method with a series of strong baselines to evaluate performance on the few-shot ACSA task.

- **Question-Driven** (Sun et al., 2019): For an aspect, it designs the corresponding question prompt to guide a large language model (e.g., BERT) to identify sentiment polarity towards the aspect. For example, the prompt is “*The polarity of the aspect safety is positive*”, and then the large language model outputs a probability value of *yes* as the matching score to determine if the sentiment of *safety* is positive. Although the method achieves impressive performance, it heavily relies on the quality of prompts. It is hard to find a great prompt to obtain the best performance.
- **MIMLLN** (Li et al., 2020): In a sentence, it first extracts some aspect-associated words to depict the context of a specific aspect. Then, the method combines the sentiment information of these words to predict the overall sentiment polarity towards the aspect. Though effective, it focuses on the sentiment of individual words and fails to capture the entire semantic content.
- **CapsNet** (Jiang et al., 2019): It designs a capsule-guided routing method to model the interactions between an aspect and its contexts. Specifically, the method constructs a set of capsules by linear transformation and squashing activation (Sabour et al., 2017). These capsules use aspect-associated words to construct a sentiment matrix to learn some sentiment knowledge for a specific aspect. Then, the method utilizes the sentiment matrix to learn the relationship between an

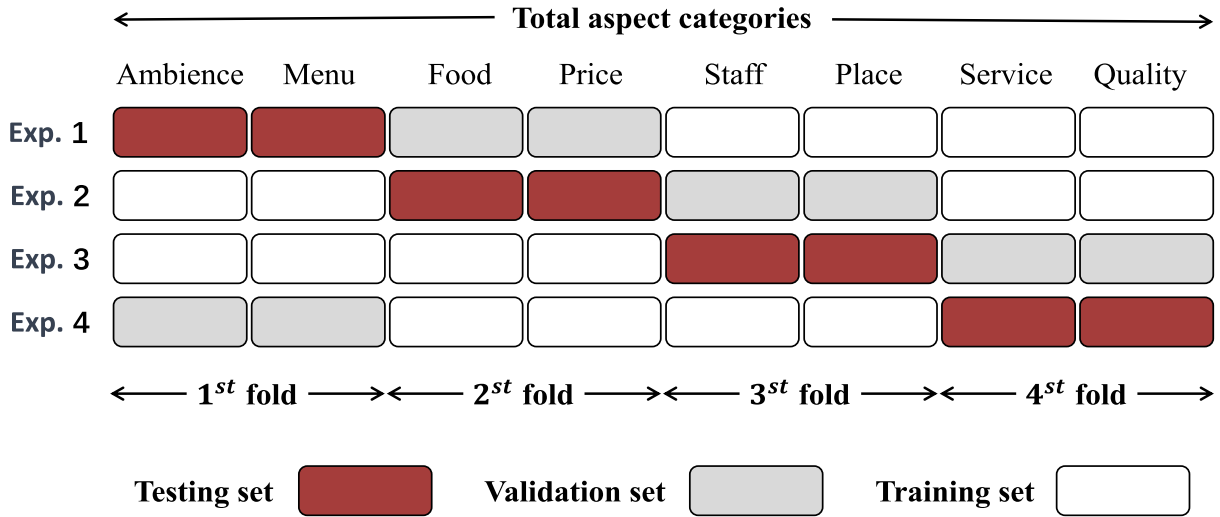


Figure 3: The schematic diagram of four-fold cross-validation.

aspect and its contexts for predicting the sentiment label of the aspect.

- Relation Network** (Sung et al., 2018): In the meta-learning formulation, a neural network computes the similarity scores between each query sample and all support samples. The similarity score represents the relation strength between the query sample and different support samples. Therefore, the method leverages the support label that exhibits the highest similarity to the query to deduce the label of a query sample.
- Induction Network** (Geng et al., 2019): It performs a matrix transformation on support samples to generate a class embedding for each sentiment label. Then, a neural tensor network (Geng et al., 2017) computes the similarity scores between each query and all class embeddings to determine which class matches the query.
- MTM** (Deng et al., 2020): It designs a meta-pretraining strategy for a large language model (e.g., BERT) to learn task-agnostic general features that extract linguistic properties to benefit downstream few-shot learning tasks. Then, task-specific parameters are fine-tuned on the large language model for the few-shot ACSA task to enable predictions aligning with its specific requirements.
- AFML** (Liang et al., 2023): It uses an existing knowledge-based method (Liang et al., 2021) to collect highly aspect-associated words from an external knowledge source (e.g., SenticNet [Cambria et al., 2020]). Then, it constructs two auxiliary sentences by masking aspect-associated words and masking non-aspect words in the original sentence. Finally, it combines these two auxiliary sentences and the original sentence to enhance the features of a specific aspect and highlight the significant contextual sentiment clues of the specific aspect to promote the sentiment prediction of the aspect.
- T5** (Raffel et al., 2020): It adopts an encoder-decoder architecture, where the few-shot ACSA task could be formulated as a text-to-text problem. Specifically, the encoder part encodes a sentence into hidden states, and the decoder part takes the encoder outputs and a specific aspect as inputs to identify the sentiment polarity of the aspect. In experiments, we use T5-base to evaluate the performance of T5 for the few-shot ACSA task.
- MetaAdapt** (Yue et al., 2023): Based on meta-learning, it proposes a few-shot domain adaptation method. The method divides a dataset into a source domain and a target domain, and it constructs the support set in the source domain and the query set in the target domain. Then, it leverages the support set to train the model to obtain gradients and evaluates the model on the query set to get second-order gradients w.r.t. the

Models	Rest I				Rest II			
	3-way(%)		2-way(%)		3-way(%)		2-way(%)	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Relation Network (Sung et al., 2018)	53.25	70.36	70.72	87.13	58.32	73.50	78.72	82.25
MTM (Deng et al., 2020)	53.93	57.15	66.23	69.10	63.12	63.71	72.13	73.87
Induction Network (Geng et al., 2019)	72.03	74.75	82.77	85.96	76.53	78.17	82.70	83.55
MIMLLN (Li et al., 2020)	74.63	74.07	87.19	87.32	79.21	78.46	81.26	81.97
Question-Driven (Sun et al., 2019)	74.66	74.79	87.52	86.83	78.69	79.84	82.63	83.31
CapsNet (Jiang et al., 2019)	75.18	73.92	87.10	87.15	78.72	80.11	81.57	82.18
MetaAdapt (Yue et al., 2023)	65.06	74.74	87.69	86.42	70.20	79.44	82.68	83.58
T5 (Raffel et al., 2020)	77.01	<u>78.15</u>	81.61	85.85	81.26	<u>82.36</u>	<u>85.48</u>	87.57
AFML (Liang et al., 2023)	<u>77.13</u>	<u>77.53</u>	<u>88.12</u>	88.79	<u>81.56</u>	81.95	83.89	84.15
Our method	78.17	78.64	88.75	<u>87.74</u>	81.58	84.18	87.22	<u>86.85</u>

Table 3: Comparison of accuracy on Rest I and Rest II.

Models	Lap				Mams			
	3-way(%)		2-way(%)		3-way(%)		2-way(%)	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Relation Network (Sung et al., 2018)	57.15	69.03	80.80	85.91	37.20	36.91	58.32	62.19
MTM (Deng et al., 2020)	51.99	53.22	66.15	68.19	37.58	36.26	58.33	57.90
Induction Network (Geng et al., 2019)	70.01	70.53	87.18	86.43	38.20	35.46	62.75	59.31
MIMLLN (Li et al., 2020)	68.79	70.03	87.03	86.18	36.52	37.43	62.30	63.17
Question-Driven (Sun et al., 2019)	70.30	68.82	86.17	87.15	36.08	35.44	63.17	61.05
CapsNet (Jiang et al., 2019)	71.53	69.82	86.82	86.73	37.12	36.98	61.61	63.75
MetaAdapt (Yue et al., 2023)	60.76	69.45	87.36	87.52	42.03	43.60	62.04	64.48
T5 (Raffel et al., 2020)	<u>74.72</u>	<u>75.12</u>	89.18	89.42	<u>46.09</u>	47.38	<u>68.82</u>	72.31
AFML (Liang et al., 2023)	72.96	73.80	88.17	<u>88.67</u>	40.07	40.35	65.57	66.30
Our method	76.05	75.51	<u>88.21</u>	87.75	47.66	<u>45.86</u>	69.77	<u>68.56</u>

Table 4: Comparison of accuracy on Lap and Mams.

original parameters. Additionally, it computes the similarity between the original and second-order gradients to select more ‘informative’ support samples. These selected support samples are used to reweight the support set to optimize the model performance in the query set. Therefore, the model can optimally adapt to the target distribution with the provided source domain knowledge.

5 Analysis & Discussion

5.1 Overall Performance

We conduct extensive experiments with 3/2-way and 1/5-shot settings on Rest I, Rest II, Lap, and Mams datasets. The results are reported in Tables 3, 4, 5 and 6, where the best scores are highlighted in bold, and the runner-up scores are marked by underline, with the following observations.

(1) Overall, the proposed method outperforms most baselines. Additionally, we also observe that two strong baselines, AFML and T5, achieve competitive results, but their overall performance is much worse than our proposed method. Specifically, in terms of accuracy, the proposed method improves upon the strong baseline AFML up to an average of 0.43%, 2.07%, 0.98%, and 4.89% on Rest I, Rest II, Lap, and Mams, respectively. Compared to T5, the proposed method achieves an average of 2.67% and 0.79% accuracy improvements on Rest I and Rest II, respectively. Although T5 obtains competitive performance on Lap and Mams, it only outperforms our proposed method by a margin of 0.23% and 0.68% in accuracy. More specifically, as for accuracy, in Table 3 and Table 4, T5 obtains convincing performance on five scenarios due to the advantage of abundant pre-trained knowledge in its encoder-decoder architecture, but it has worse results than our proposed method on the other 11 scenarios. Therefore,

Models	Rest I				Rest II			
	3-way(%)		2-way(%)		3-way(%)		2-way(%)	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Relation Network (Sung et al., 2018)	52.19	60.75	67.28	83.41	52.87	61.34	71.34	78.95
MTM (Deng et al., 2020)	53.83	53.79	63.54	65.19	61.14	60.05	70.47	71.45
Induction Network (Geng et al., 2019)	60.89	60.51	79.62	81.34	63.08	62.41	80.49	80.64
MIMLLN (Li et al., 2020)	61.24	61.53	82.08	<u>83.44</u>	64.21	63.20	78.84	79.51
Question-Driven (Sun et al., 2019)	62.69	62.52	<u>83.93</u>	83.40	64.18	<u>64.61</u>	80.03	80.15
CapsNet (Jiang et al., 2019)	60.84	60.82	82.44	83.01	65.78	<u>64.29</u>	81.22	79.85
MetaAdapt (Yue et al., 2023)	58.36	61.22	82.42	82.20	62.02	62.43	81.40	82.20
T5 (Raffel et al., 2020)	55.12	<u>63.15</u>	75.55	80.14	59.12	61.33	<u>84.25</u>	85.05
AFML (Liang et al., 2023)	<u>64.05</u>	62.87	74.53	74.68	<u>66.19</u>	63.58	81.74	81.28
Our method	64.49	63.30	86.06	85.16	67.62	66.51	85.03	<u>84.55</u>

Table 5: Comparison of F1 score on Rest I and Rest II.

Models	Lap				Mams			
	3-way(%)		2-way(%)		3-way(%)		2-way(%)	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Relation Network (Sung et al., 2018)	49.10	53.35	80.61	80.59	34.79	35.36	56.00	60.79
MTM (Deng et al., 2020)	50.11	51.47	62.23	64.69	36.96	35.21	52.23	50.65
Induction Network (Geng et al., 2019)	54.67	54.79	83.92	84.56	37.15	34.54	60.08	58.78
MIMLLN (Li et al., 2020)	53.71	53.66	84.59	83.87	36.03	36.92	59.41	60.59
Question-Driven (Sun et al., 2019)	<u>54.80</u>	54.43	84.75	84.21	35.79	34.66	60.37	60.13
CapsNet (Jiang et al., 2019)	54.30	53.26	83.53	83.39	35.92	35.12	58.63	62.65
MetaAdapt (Yue et al., 2023)	50.28	53.34	85.72	86.35	35.64	35.57	60.31	63.70
T5 (Raffel et al., 2020)	53.63	<u>55.37</u>	87.43	88.12	<u>38.53</u>	41.46	66.48	70.53
AFML (Liang et al., 2023)	54.75	52.06	85.92	86.27	38.46	34.09	64.36	65.33
Our method	58.91	56.49	86.15	<u>87.52</u>	40.84	<u>38.82</u>	66.54	<u>66.48</u>

Table 6: Comparison of F1 score on Lap and Mams.

our proposed method performs better than T5 for the few-shot ACSA task overall. Regarding F1 score, the proposed method improves upon AFML by 0.23%–11.53% on Rest I, Rest II, Lap, and Mams. The proposed method improves upon T5 by an average of 6.26%, 3.49%, and 1.13% F1 scores on Rest I, Rest II, and Lap, respectively. Although T5 obtains competitive results for the 2-way setting on Lap, its average performance is much worse due to the low F1 score for the 3-way setting. As for F1 score, although T5 also obtains convincing performance on five scenarios in Table 5 and Table 6, our proposed method still outperforms the other 11 scenarios. The results demonstrate the effectiveness of the proposed method for the few-shot ACSA task. The proposed method learns the similarity and diversity relations among support and query samples to alleviate irrelevant sentiment noise and effectively predict query labels by exploiting intra-cluster commonality and inter-cluster uniqueness.

(2) For all mentioned methods, the performance of the 3-way setting on Rest I is inferior to those on Rest II. This is because Rest II provides a more fine-grained aspect (i.e., entity#attribute), whereas Rest I only gives a general aspect (i.e., entity). For the 3-way setting, the proposed method surpasses the strong baseline AFML by an average improvement of 1.07% accuracy and 0.43% F1 score on Rest I and 1.12% accuracy and 2.18% F1 score on Rest II. Besides, in the 3-way setting, the proposed method surpasses the strong baseline T5 by an average improvement of 0.82% accuracy and 4.76% F1 score on Rest I and 1.07% accuracy and 6.84% F1 score on Rest II. The results indicate that our proposed method performs better when given fine-grained aspects.

(3) Compared to generative model T5, the proposed method achieves accuracy/F1 improvements on Rest I, Rest II, Lap, and Mams. In experiments, the proposed method is based on the BERT-base model and freezes half of the

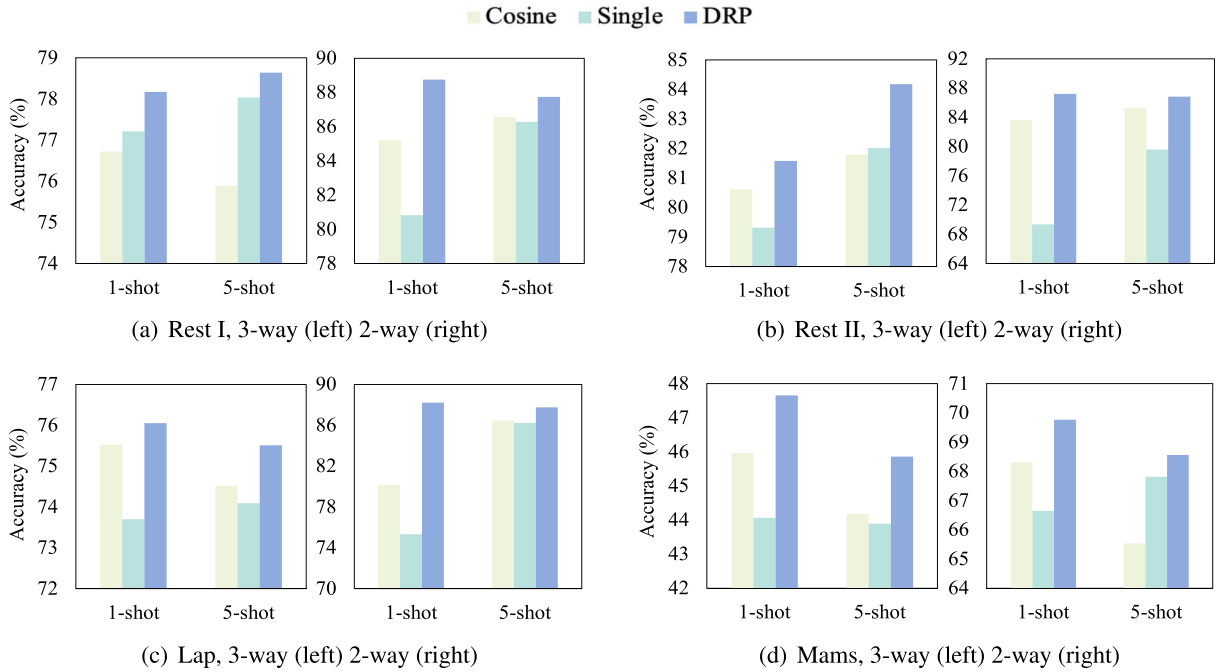


Figure 4: The impact of DRP on Rest I, Rest II, Lap, and Mams datasets based on the accuracy metric.

parameters. However, the T5 model has large amounts of parameters due to the advantage of its encoder-decoder architecture. To ensure a fair comparison, we freeze the encoder of the strong baseline T5 (i.e., T5-base) to compare performance with our proposed method. Although T5 obtains some competitive results due to its abundant pre-trained knowledge, it performs worse than our proposed method in most experiments. Significantly, our proposed method improves upon T5 by 10.51% at most and an average of 6.26% F1 score on Rest I. Additionally, our proposed method achieves an average of 6.36% F1 score improvements in the 3-way 1-shot setting. The results denote that T5 performs poorly on the few-shot ACSA task due to irrelevant sentiment noise in the above complex few-shot scenario. In short, our proposed method alleviates irrelevant sentiment noise to improve performance in few-shot scenarios by exploring intra-cluster commonality and inter-cluster uniqueness.

5.2 Impact of Dual Relations Propagation

DRP is used to enhance similarity and diversity relations among samples by exploiting intra-cluster commonality and inter-cluster uniqueness. To verify the impact of DRP, we design two cases to evaluate the performance of the proposed method. The first case uses cosine distance between a node pair to replace the learning of dual

relations to evaluate the importance of dual relations propagation and aggregation. In another case, the proposed method only learns the similarity relation among samples to analyze whether dual relations are essential. Thus, we redefine the relation graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}^+)$ and use $\mathcal{E}^+ = \{E_{ij}^+\}_{i,j}^M$ instead of dual edges to evaluate the proposed method.

The experimental results are presented in Figure 4 for the Rest I, Rest II, Lap, and Mams datasets. We can observe that DRP performs best in these two cases. Specifically, DRP better enhances the similarity and diversity relations among samples than simple cosine distances among those to promote query inference to improve performance. Also, DRP obtains more convincing results than single similarity relation propagation because DRP considers contrastive enhancement between similarity and diversity relations. Furthermore, most results of the cosine distance are better than the single similarity relation propagation. This indicates that single similarity relation propagation is weak in few-shot scenarios. Besides, the performances of all methods are worse in the 3-way setting than those in the 2-way setting due to sentiment complicity. However, DRP can still achieve the best results in Figure 4. The experimental results verify the effectiveness of DRP in the proposed method, revealing it is vital to guarantee good performance.

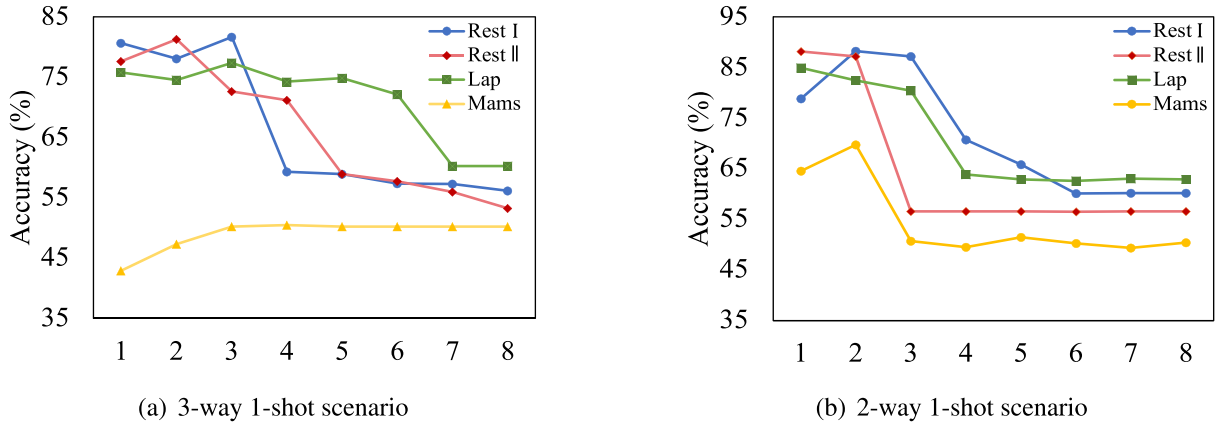


Figure 5: The impact of relation propagation layers.

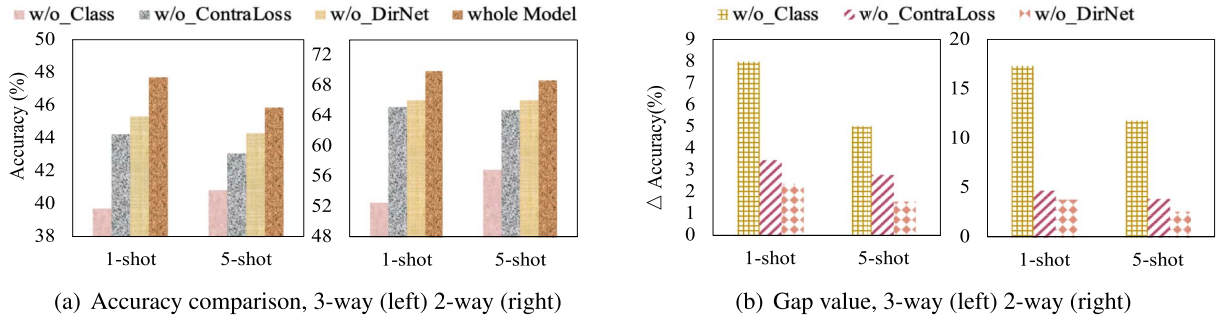


Figure 6: Experimental results of ablation study.

5.3 Impact of Propagation Layer

To investigate the impact of the DRP layer, we evaluate the proposed method with one to eight layers on Rest I, Rest II, Lap, and Mams. Figure 5 shows the performance of the proposed method with increasing the number of layers in 3-way 1-shot and 2-way 1-shot scenarios. In terms of a 3-way 1-shot scenario, DRP with two layers obtains the best results on Rest II while DRP with three layers performs best on Rest I, Lap, and Mams. In terms of a 2-way 1-shot scenario, DRP with two layers obtains the best results on Rest I and Mams. The results indicate that DRP plays a positive effect during relation propagation with increasing layers, but excessive layers result in low performance due to over-fitting. In short, the results demonstrate the effectiveness of DRP within limited layers.

5.4 Ablation Study

To investigate the significance of the proposed method, we conduct an ablation study on the most competitive Mams dataset to compare perfor-

mance. Due to the complexity of Mams, we can observe convincing differences in ablation experiments. Experimental results are shown in Figure 6, where the comparison results are presented in Figure 6a, and the gap values are reported in Figure 6b, with the following observations.

(1) In the “w/o_Class”, we remove the relation transformation from support-query to class-query and only use support-query relations to induce the labels of query samples. The performance is significantly reduced when removing the class-query relation transformation mechanism. This fact indicates that modeling the relations between a query and classes can promote query inference. Therefore, the class-query relation transformation plays an important role in the performance of the proposed method.

(2) In the “w/o_ContraLoss”, we remove the training objective mentioned in Section 3.6 and use a cross-entropy loss instead. When the proposed training objective is removed, the performance drops considerably. The negative effect suggests that the proposed training objective has a positive role in capturing dual relations features. The proposed training objective promotes the

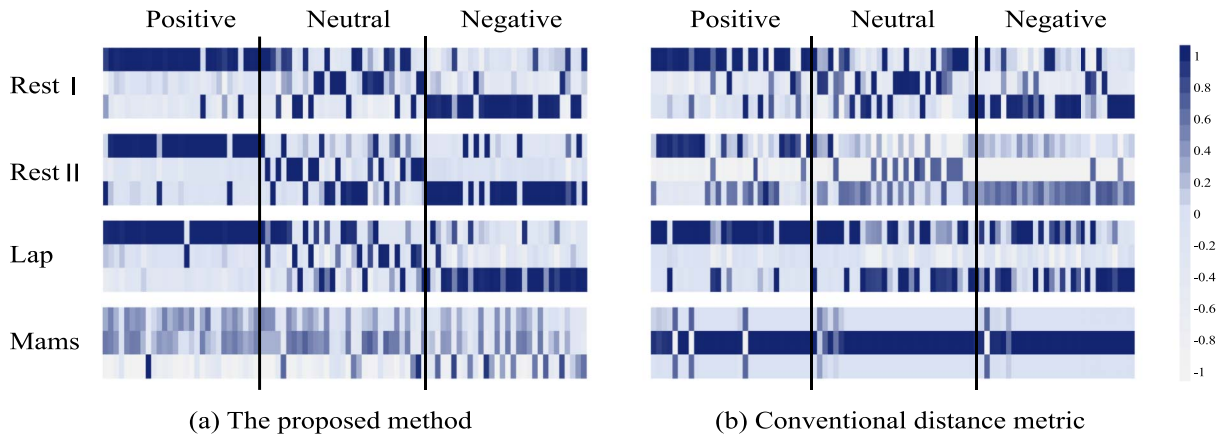


Figure 7: Visualization of relation strength from sentiment classes to query samples. From top to bottom, the illustration shows 3-way 5-shot scenarios on Rest I, Rest II, Lap, and Mams datasets. Each scenario includes 90 query samples, with positive samples for the first thirty, neutral samples for the middle thirty, and negative samples for the last thirty. Briefly, the scenario of each dataset is described by a heat map with a size of $3 * 90$. The heat map shows the relation strength among positive, neutral, and negative classes and samples with different sentiments. Dark color denotes higher relation scores, while light color denotes lower those.

model to explore intra-cluster commonality and inter-cluster uniqueness to learn discriminative dual relations for query inference. Therefore, the training objective has considerable importance.

(3) In the “w/o_DirNet”, we replace the learning of diversity relation with a simple mechanism. Specifically, we set the strongest relation score to 1 and replace $e_{ij}^{-(l)}$ with $(1 - e_{ij}^{+(l)})$ as the diversity score to analyze the necessity of the diversity network in Equation 15. For example, if the similarity score is 0.7, we set 0.3 as its corresponding diversity score, i.e., there are 0.7 similarity relation and 0.3 diversity relation between a sample pair. Experimental results demonstrate that the diversity network provides more supplementary information and encourages DRP to learn more robust dual relations among samples. Therefore, the learning of the diversity relation is essential for the few-shot ACSA task.

Obviously, the absence of any method part can decrease performance. In short, the whole model consistently surpasses all ablation studies and achieves the best performance.

5.5 Relation Strength Visualization

The relation strength between query samples and sentiment classes is visualized to verify the performance of DRP compared with conventional distance metrics (e.g., cosine distance). Specifically, we use similarity relation score and cosine distance to draw heat maps in Figure 7. The experimental results are conducted on Rest I, Rest II,

Lap, and Mams datasets. It is easy to find the following observations.

(1) For Rest I, Rest II, and Lap, the proposed method significantly improves positive and negative class results. Regarding the neutral class, these two methods are weak in the relation drawing between query samples and the neutral class because many queries with neutral sentiment are predicted to be positive and negative. However, the proposed method performs better than the conventional distance metric. The proposed method learns discriminative relation features to improve performance on the few-shot ACSA task by modeling the dual relations among samples.

(2) Mams provides a more complex scenario where each sentence contains many aspects with different sentiment polarities. Inevitably, there are overlapping distributions of aspect embeddings caused by irrelevant sentiment noise among sentences with multiple sentiment aspects. Therefore, the conventional distance method mainly classifies everything into the neutral class, failing to identify sentiment features. Relatively, the proposed method learns valuable features to classify different sentiment classes compared with the conventional distance method, but it still has a weak relation strength between query samples and sentiment classes. Therefore, in the following work, we could focus on refining the feature extraction and exploring additional domain knowledge to enhance the relation strength between query samples and sentiment classes for query inference.

6 Conclusion & Future Work

We propose an effective metric-free method for the few-shot ACSA task, which explicitly models the associated relations among aspects of query and support samples, addressing the passive effect of overlapping distributions caused by irrelevant sentiment noise in aspect distributions. Specifically, the proposed method designs a fully connected relation graph to model the dual relations (similarity and diversity) among support and query nodes in the embedding space. With the relation graph, the proposed method uses the dual relations among nodes to explore intra-cluster commonality and inter-cluster uniqueness to alleviate irrelevant sentiment noise and enhance aspect features, eliminating the passive effect of overlapping distributions. Additionally, the dual relations are transformed from support-query to class-query to guide query inference by learning sentiment class knowledge from the relation graph. Experiments show that the proposed method outperforms strong baselines and obtains significant performance.

The proposed method is not limited to few-shot ACSA, and it can be applied to more complex tasks, e.g., fake news detection, text classification, and intention detection, since it could better enhance semantic textual similarity and diversity with ground truth texts. Therefore, we will extend our method to these tasks in follow-up work.

7 Limitations

The proposed method obtains convincing performance compared with baselines but still has a few limitations.

(1) The neutral classification for the few-shot ACSA task is still a considerable challenge. The neutral classification lacks opinion sentiment contexts, making it susceptible to irrelevant sentiment noise, resulting in wrong classifications, as shown in Figure 7. Therefore, we want to attract more researchers and developers to pay more attention to the challenge of neutral classification.

(2) We follow the meta-learning formulation to perform few-shot ACSA. For meta-learning, the meta-task structure could generalize experiences from seen aspects to newly encountered aspects but requires annotating a few samples to construct the support set from the newly encountered aspects for query inference. Therefore, we could focus

on reducing shots per class in the sample set, eliminating the burden of data annotation.

Acknowledgments

The authors are grateful for helpful comments from the anonymous reviewers and the ACL action editor. This work is supported by the National Key Research and Development Program of China (no. 2021ZD0111202), the National Natural Science Foundation of China (no. 62176005), and the Art Project of the National Social Science Fund of China (no. 2022CC02195).

References

- Sajad Ahmadian, Milad Ahmadian, and Mahdi Jalili. 2022. A deep learning based trust-and tag-aware recommender system. *Neurocomputing*, 488:557–571. <https://doi.org/10.1016/j.neucom.2021.11.064>
- Mahmoud Assran, Mathilde Caron, Ishan Misra, Piotr Bojanowski, Florian Bordes, Pascal Vincent, Armand Joulin, Mike Rabbat, and Nicolas Ballas. 2022. Masked Siamese networks for label-efficient learning. In *Proceedings of the European Conference on Computer Vision*, pages 456–473. https://doi.org/10.1007/978-3-031-19821-2_26
- Hongjie Cai, Rui Xia, and Jianfei Yu. 2021. Aspect-category-opinion-sentiment quadruple extraction with implicit aspects and opinions. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pages 340–350. <https://doi.org/10.18653/v1/2021.acl-long.29>
- Erik Cambria, Yang Li, Frank Z. Xing, Soujanya Poria, and Kenneth Kwok. 2020. Senticnet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 105–114. <https://doi.org/10.1145/3340531.3412003>
- Chenhua Chen, Zhiyang Teng, Zhongqing Wang, and Yue Zhang. 2022a. Discrete opinion tree

- induction for aspect-based sentiment analysis. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pages 2051–2064. <https://doi.org/10.18653/v1/2022.acl-long.145>
- Junfan Chen, Richong Zhang, Yongyi Mao, and Jie Xu. 2022b. Contrastnet: A contrastive learning framework for few-shot text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 10492–10500. <https://doi.org/10.1609/aaai.v36i10.21292>
- Lisong Chen, Peilin Zhou, and Yuexian Zou. 2022c. Joint multiple intent detection and slot filling via self-distillation. In *ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7612–7616. <https://doi.org/10.1109/ICASSP43922.2022.9747843>
- Xiudi Chen, Hui Wu, and Xiaodong Shi. 2023. Consistent prototype learning for few-shot continual relation extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 7409–7422. <https://doi.org/10.18653/v1/2023.acl-long.409>
- Zhihua Cui, Xianghua Xu, Xue Fei, Xingjuan Cai, Yang Cao, Wensheng Zhang, and Jinjun Chen. 2020. Personalized recommendation system based on collaborative filtering for iot scenarios. *IEEE Transactions on Services Computing*, 13(4):685–695. <https://doi.org/10.1109/TSC.2020.2964552>
- Shumin Deng, Ningyu Zhang, Zhanlin Sun, Jiaoyan Chen, and Huajun Chen. 2020. When low resource NLP meets unsupervised language model: Meta-pretraining then meta-learning for few-shot text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 13773–13774. <https://doi.org/10.1609/aaai.v34i10.7158>
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.
- Xiao Ding, Yue Zhang, Ting Liu, and Junwen Duan. 2015. Deep learning for event-driven stock prediction. In *Proceedings of the 24th International Conference on Artificial Intelligence July*, pages 2327–2333. <https://doi.org/10.5555/2832415.2832572>
- Thomas Efland and Michael Collins. 2023. Improving low-resource cross-lingual parsing with expected statistic regularization. *Transactions of the Association for Computational Linguistics*, 11:122–138. https://doi.org/10.1162/tacl_a_00537
- Jinyuan Fang, Xiaobin Wang, Zaiqiao Meng, Pengjun Xie, Fei Huang, and Yong Jiang. 2023. MANNER: A variational memory-augmented model for cross domain few-shot named entity recognition. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 4261–4276. <https://doi.org/10.18653/v1/2023.acl-long.234>
- Honghao Gao, Jiadong Huang, Yuan Tao, Walayat Hussain, and Yuzhe Huang. 2022. The joint method of triple attention and novel loss function for entity relation extraction in small data-driven computational social systems. *IEEE Transactions on Computational Social Systems*, 9(6):1725–1735. <https://doi.org/10.1109/TCSS.2022.3178416>
- Ruiying Geng, Ping Jian, Yingxue Zhang, and Heyan Huang. 2017. Implicit discourse relation identification based on tree structure neural network. In *2017 International Conference on Asian Language Processing (IALP)*, pages 334–337. <https://doi.org/10.1109/IALP.2017.8300611>
- Ruiying Geng, Binhua Li, Yongbin Li, Xiaodan Zhu, Ping Jian, and Jian Sun. 2019. Induction networks for few-shot text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 3895–3904. <https://doi.org/10.18653/v1/D19-1403>
- Ehsan Hosseini-Asl, Wenhao Liu, and Caiming Xiong. 2022. A generative language model for few-shot aspect-based sentiment analysis. In *Findings of the 2022 Conference of the North*

- American Chapter of the Association for Computational Linguistics: Human Language Technologies*. <https://doi.org/10.18653/v1/2022.findings-naacl.58>
- Yutai Hou, Yongkui Lai, Yushan Wu, Wanxiang Che, and Ting Liu. 2021. Few-shot learning for multi-label intent detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 13036–13044. <https://doi.org/10.1609/aaai.v35i14.17541>
- Mengting Hu, Shiwan Zhao, Honglei Guo, Chao Xue, Hang Gao, Tiegang Gao, Renhong Cheng, and Zhong Su. 2021. Multi-label few-shot learning for aspect category detection. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*. <https://doi.org/10.18653/v1/2021.acl-long.495>
- Shell Xu Hu, Da Li, Jan Stühmer, Minyoung Kim, and Timothy M. Hospedales. 2022. Pushing the limits of simple pipelines for few-shot learning: External data and fine-tuning make a difference.
- Chao Huang, Zhangjie Cao, Yunbo Wang, Jianmin Wang, and Mingsheng Long. 2021. Metasets: Meta-learning on point sets for generalizable representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. <https://doi.org/10.1109/CVPR46437.2021.00875>
- Dietmar Jannach, Ahtsham Manzoor, Wanling Cai, and Li Chen. 2021. A survey on conversational recommender systems. *ACM Computing Surveys (CSUR)*, 54(5):1–36. <https://doi.org/10.1145/3453154>
- Qingnan Jiang, Lei Chen, Ruifeng Xu, Xiang Ao, and Min Yang. 2019. A challenge dataset and effective models for aspect-based sentiment analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pages 6280–6285. <https://doi.org/10.18653/v1/D19-1654>
- Jaejun Lee, Raphael Tang, and Jimmy Lin. 2019a. What would Elsa do? Freezing layers during transformer fine-tuning. *arXiv preprint arXiv:1911.03090v1*. <https://doi.org/10.48550/arXiv.1911.03090>
- Kwonjoon Lee, Subhansu Maji, Avinash Ravichandran, and Stefano Soatto. 2019b. Meta-learning with differentiable convex optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10657–10665.
- Jia Li, Chongyang Tao, Huang Hu, Can Xu, Yining Chen, and Daxin Jiang. 2022a. Unsupervised cross-domain adaptation for response selection using self-supervised and adversarial training. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 562–570. <https://doi.org/10.1145/3488560.3498404>
- Jia Li, Yuyuan Zhao, Zhi Jin, Ge Li, Tao Shen, Zhengwei Tao, and Chongyang Tao. 2022b. SK2: Integrating implicit sentiment knowledge and explicit syntax knowledge for aspect-based sentiment analysis. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pages 1114–1123. <https://doi.org/10.1145/3511808.3557452>
- Peng Li, Tianxiang Sun, Qiong Tang, Hang Yan, Yuanbin Wu, Xuanjing Huang, and Xipeng Qiu. 2023. CodeIE: Large code generation models are better few-shot information extractors. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, pages 15339–15353. <https://doi.org/10.48550/arXiv.2305.05711>
- Ruifan Li, Hao Chen, Fangxiang Feng, Zhanyu Ma, Xiaojie Wang, and Eduard Hovy. 2021a. Dual graph convolutional networks for aspect-based sentiment analysis. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pages 6319–6329. <https://doi.org/10.18653/v1/2021.acl-long.494>
- Yuzhi Li, Siwei Li, Jing Hu, Yuanyuan Zhang, Yuchen Du, Xijiang Han, Xi Liu, and Ping Xu. 2021b. Hollow feco-fecop@c nanocubes embedded in nitrogen-doped carbon nanocages for efficient overall water splitting. *Journal of Energy Chemistry*, 53:1–8. <https://doi.org/10.1016/j.jechem.2020.05.012>

- Yuncong Li, Cunxiang Yin, Sheng-hua Zhong, and Xu Pan. 2020. Multi-instance multi-label learning networks for aspect-category sentiment analysis. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pages 3550–3560. <https://doi.org/10.18653/v1/2020.emnlp-main.287>
- Bin Liang, Xiang Li, Lin Gui, Yonghao Fu, Yulan He, Min Yang, and Ruifeng Xu. 2023. Few-shot aspect category sentiment analysis via meta-learning. *ACM Transactions on Information Systems*, 41(1):1–31. <https://doi.org/10.1145/3529954>
- Bin Liang, Rongdi Yin, Jiachen Du, Lin Gui, Yulan He, Min Yang, and Ruifeng Xu. 2021. Embedding refinement framework for targeted aspect-based sentiment analysis. *IEEE Transactions on Affective Computing*.
- Han Liu, Feng Zhang, Xiaotong Zhang, Siyang Zhao, Junjie Sun, Hong Yu, and Xianchao Zhang. 2022a. Label-enhanced prototypical network with contrastive learning for multi-label few-shot aspect category detection. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1079–1087. <https://doi.org/10.1145/3534678.3539340>
- MeiZhen Liu, FengYu Zhou, Ke Chen, and Yang Zhao. 2021. Co-attention networks based on aspect and context for aspect-level sentiment analysis. *Knowledge-Based Systems*, 217:106810. <https://doi.org/10.1016/j.knosys.2021.106810>
- Qian Liu, Zhiqiang Gao, Bing Liu, and Yuanlin Zhang. 2015. Automated rule selection for aspect extraction in opinion mining. In *Proceedings of International Joint Conference on Artificial Intelligence*, pages 1291–1297.
- Yang Liu, Weifeng Zhang, Chao Xiang, Tu Zheng, Deng Cai, and Xiaofei He. 2022b. Learning to affiliate: Mutual centralized learning for few-shot classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14391–14400. <https://doi.org/10.1109/CVPR52688.2022.01401>
- Yaojie Lu, Qing Liu, Dai Dai, and Xinyan Xiao. 2022. Unified structure generation for universal information extraction. In *Proceedings of the Conference on Association for Computational Linguistics*. <https://doi.org/10.18653/v1/2022.acl-long.395>
- Hui Lv, Chen Chen, Zhen Cui, Chunyan Xu, Yong Li, and Jian Yang. 2021. Learning normal dynamics in videos with meta prototype network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15425–15434. <https://doi.org/10.1109/CVPR46437.2021.01517>
- Ruotian Ma, Zhang Lin, Xuanting Chen, Xin Zhou, Junzhe Wang, Tao Gui, Qi Zhang, Xiang Gao, and Yunwen Chen. 2023. Coarse-to-fine few-shot learning for named entity recognition. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 4115–4129. <https://doi.org/10.18653/v1/2023.findings-acl.253>
- Cheng Ouyang, Carlo Biffi, Chen Chen, Turkey Kart, Huaqi Qiu, and Daniel Rueckert. 2022. Self-supervised learning for few-shot medical image segmentation. *IEEE Transactions on Medical Imaging*, 41(7):1837–1848. <https://doi.org/10.1109/TMI.2022.3150682>, PubMed: 35139014
- Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, and Wei Li. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*.
- Sara Sabour, Nicholas Frosst, and Geoffrey E. Hinton. 2017. Dynamic routing between capsules. *Advances in Neural Information Processing Systems*, 30.
- Ronald Seoh, Ian Birle, Mrinal Tak, Haw-Shiuan Chang, Brian Pinette, and Alfred Hough. 2021. Open aspect target sentiment classification with natural language prompts. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6311–6322. <https://doi.org/10.18653/v1/2021.emnlp-main.509>
- Jinsong Su, Jialong Tang, Hui Jiang, Ziyao Lu, Yubin Ge, Linfeng Song, Deyi Xiong, Le Sun,

- and Jiebo Luo. 2021. Enhanced aspect-based sentiment analysis models with progressive self-supervised attention learning. *Artificial Intelligence*, 296:103477. <https://doi.org/10.1016/j.artint.2021.103477>
- Jianlin Su, Mingren Zhu, Ahmed Murtadha, Shengfeng Pan, Bo Wen, and Yunfeng Liu. 2022. Zlpr: A novel loss for multi-label classification. *arXiv preprint arXiv:2208.02955v1*. <https://doi.org/10.48550/arXiv.2208.02955>
- Chi Sun, Luyao Huang, and Xipeng Qiu. 2019. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 380–385.
- Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip H. S. Torr, and Timothy M. Hospedales. 2018. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1199–1208. <https://doi.org/10.1109/CVPR.2018.00131>
- Yue Tan, Guodong Long, Lu Liu, Tianyi Zhou, Qinghua Lu, Jing Jiang, and Chengqi Zhang. 2022. FedProto: Federated prototype learning across heterogeneous clients. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8432–8440. <https://doi.org/10.1609/aaai.v36i8.20819>
- Yuanhe Tian, Guimin Chen, and Yan Song. 2021. Aspect-based sentiment analysis with type-aware graph convolutional networks and layer ensemble. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 2910–2922. <https://doi.org/10.18653/v1/2021.naacl-main.231>
- Munkhdalai Tsendsuren and Yu Hong. 2017. Meta networks. In *International Conference on Machine Learning*, pages 2554–2563.
- Runzhong Wang, Junchi Yan, and Xiaokang Yang. 2021. Neural graph matching network: Learning lawler’s quadratic assignment problem with extension to hypergraph and multiple-graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9):5261–5279. <https://doi.org/10.1109/TPAMI.2021.3078053> PubMed: 33961550
- Chao Wu, Qingyu Xiong, Zhengyi Yang, Min Gao, Qiude Li, Yang Yu, Kaige Wang, and Qiwu Zhu. 2021. Residual attention and other aspects module for aspect-based sentiment analysis. *Neurocomputing*, 435:42–52. <https://doi.org/10.1016/j.neucom.2021.01.019>
- Luwei Xiao, Yun Xue, Hua Wang, Xiaohui Hu, Donghong Gu, and Yongsheng Zhu. 2022. Exploring fine-grained syntactic information for aspect-based sentiment classification with dual graph neural networks. *Neurocomputing*, 471:48–59. <https://doi.org/10.1016/j.neucom.2021.10.091>
- Zeguan Xiao, Jiarun Wu, Qingliang Chen, and Congjian Deng. 2021. BERT4GCN: Using BERT intermediate layers to augment GCN for aspect-based sentiment classification. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 9193–9200. <https://doi.org/10.18653/v1/2021.emnlp-main.724>
- Yuanyuan Xu, Zeng Yang, Linhai Zhang, Deyu Zhou, Tiandeng Wu, and Rong Zhou. 2023. Focusing, bridging and prompting for few-shot nested named entity recognition. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 2621–2637. <https://doi.org/10.18653/v1/2023.findings-acl.164>
- Zhuo Yang, Yufei Han, Guoxian Yu, Qiang Yang, and Xiangliang Zhang. 2020. Prototypical networks for multi-label learning. In *Proceedings of the International Conference Association for the Advancement of Artificial Intelligence*. <https://doi.org/10.48550/arXiv.1911.07203>
- Tianyuan Yu, Sen He, Yi-Zhe Song, and Tao Xiang. 2022. Hybrid graph neural networks for few-shot learning. In *Proceedings of the AAAI conference on artificial intelligence*,

- volume 36, pages 3179–3187. <https://doi.org/10.1609/aaai.v36i3.20226>
- Zhenrui Yue, Huimin Zeng, Yang Zhang, Lanyu Shang, and Dong Wang. 2023. MetaAdapt: Domain adaptive few-shot misinformation detection via meta learning. In *Proceedings of the 61th Annual Meeting of the Association for Computational Linguistics*. <https://doi.org/10.48550/arXiv.2305.12692>
- Baoquan Zhang, Xutao Li, Shanshan Feng, Yunming Ye, and Rui Ye. 2022a. MetaNODE: Prototype optimization as a neural ODE for few-shot learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 9014–9021. <https://doi.org/10.1609/aaai.v36i8.20885>
- Yi Zhang, Tao Ge, and Xu Sun. 2020. Parallel data augmentation for formality style transfer. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3221–3228. <https://doi.org/10.48550/arXiv.2005.07522>, <https://doi.org/10.18653/v1/2020.acl-main.294>
- Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Cheekong Lee. 2022b. ProtGNN: Towards self-explaining graph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 9127–9135. <https://doi.org/10.1609/aaai.v36i8.20898>
- Shiman Zhao, Wei Chen, and Tengjiao Wang. 2023. Learning few-shot sample-set operations for noisy multi-label aspect category detection. In *Proceedings of International Joint Conference on Artificial Intelligence*. <https://doi.org/10.24963/ijcai.2023/589>
- Yunhua Zhou, Peiju Liu, and Xipeng Qiu. 2022. KNN-contrastive learning for out-of-domain intent classification. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, pages 5129–5141. <https://doi.org/10.18653/v1/2022.acl-long.352>