Instance Relation Learning Network with Label Knowledge Propagation for Few-shot Multi-label Intent Detection

Shiman Zhao[1,2,3] , **Shangyuan Li**[1,2,3] , **Wei Chen**[*,1,2,3] , **Tengjiao Wang**[1,2,3,4]
Jiahui Yao[1,2,3] , **Jiabin Zheng**[1,2,3] and **Kam Fai Wong**[1,4]

[1]Key Lab of High Confidence Software Technologies (MOE), School of Computer Science, Peking University, Beijing, China.
[2]Research Center for Computational Social Science, Peking University, Beijing, China.
[3]Institute of Computational Social Science, Peking University, Qingdao, China.
[4]Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong, China.
shiman.zha@gmail.com, {pekingchenwei, tjwang, isssyaojh,jiabinzheng}@pku.edu.cn

Abstract

Few-shot Multi-label Intent Detection (MID) is crucial for dialogue systems, aiming to detect multiple intents of utterances in low-resource dialogue domains. Previous studies focus on a two-stage pipeline. They first learn representations of utterances with multiple labels and then use a threshold-based strategy to identify multi-label results. However, these methods rely on representation classification and ignore instance relations, leading to error propagation. To solve the above issues, we propose a multi-label joint learning method for few-shot MID in an end-to-end manner, which constructs an instance relation learning network with label knowledge propagation to eliminate error propagation. Concretely, we learn the interaction relations between instances with class information to propagate label knowledge between a few labeled (support set) and unlabeled (query set) instances. With label knowledge propagation, the relation strength between instances directly indicates whether two utterances belong to the same intent for multi-label prediction. Besides, a dual relation-enhanced loss is developed to optimize support- and query-level relation strength to improve performance. Experiments show that we outperform strong baselines by an average of 9.54% AUC and 11.19% Macro-F1 in 1-shot scenarios.

1 Introduction

Multi-label Intent Detection (MID) [Zhang *et al.*, 2021; Pham *et al.*, 2023; Tu *et al.*, 2023] is a crucial component in dialogue systems [Zhang *et al.*, 2020; Han *et al.*, 2021], which aims to identify multiple user intents in a given utterance. For example, in the sentence "*What time is my meeting and what day?*", the MID task could detect two user intents, "*request_time*" and "*request_day*". With conversational
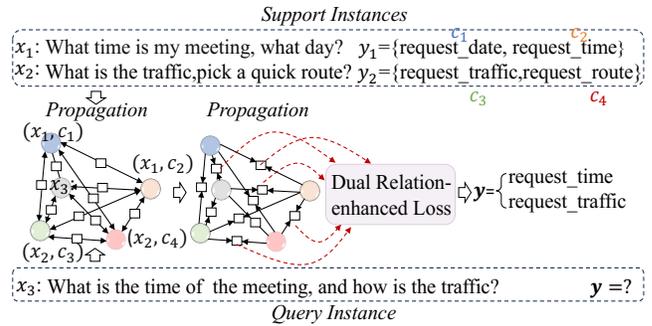
---

*Corresponding author

Figure 1: An example of the 4-way 1-shot setting.

artificial intelligence, MID has attracted widespread attention in academia and industry. Previous methods heavily rely on large amounts of labeled data, but their performances drop significantly in low-resource dialogue domains. Few-shot MID is indispensable because it can detect multiple intents of utterances when only a few labeled data are provided.

Existing methods [Hou *et al.*, 2021] mainly focus on a two-stage pipeline: (1) utterance representation learning and (2) multi-label results identification. Specifically, they first learn different representations for an utterance with multiple intents and then use these representations to obtain multi-label results through a threshold-based strategy. However, these methods heavily rely on the representation classification of utterances, leading to error propagation easily. For example, in Figure 1, they are difficult to obtain well-separated representations for intents "*request_time*" and "*request_date*" because the utterance representations of these two intents could be closer to each other due to originating from the same utterance and expressing similar label semantics, as well as "*request_traffic*" and "*request_route*". Affected by multi-label noise, they may fail to obtain reliable representations, leading to low performance. Although DCKPN [Zhang *et al.*, 2023] learns utterance representations with class information to alleviate this issue, it still underperforms in this case due to error propagation.

To solve the above issues, we are the first to propose a multi-label joint learning method for few-shot MID, which designs an instance relation learning network with label knowledge propagation to directly gain multi-label results in an end-to-end manner, addressing error propagation. Concretely, the proposed method explicitly models the relationship between intra- and inter-class instances and propagates label knowledge between instances to capture strong interaction. With label knowledge propagation, the proposed method could use the relation strength between an instance pair to indicate whether these two instances belong to the same label well for multi-label inference, eliminating the negative effect caused by error propagation. Besides, a dual relation-enhanced loss is designed to enforce support-level and query-level relation strength to further improve performance. The support-level loss enhances the relation strength between instances with the same label connected to the support instance and weakens the relation strength between those with different labels to promote label knowledge propagation. The query-level loss encourages the maximization of relation strength between a query instance and its multiple relevant support instances while minimizing relation strength between the query and its irrelevant support instances to promote multi-label prediction. Extensive experiments show that our proposed method performs well, especially an average improvement of 11.19% Macro-F1 score in the 1-shot setting. The contributions are summarized as follows:

- We propose a multi-label joint learning method for few-shot MID, which constructs an instance relation learning network with label knowledge propagation to guide query inference explicitly. The method eliminates error propagation and works well in low-resource domains.

- We model the relationship between intra- and inter-instances to directly indicate whether two utterances belong to the same intent for multi-label prediction instead of depending on representation classification.

- We design a dual relation-enhanced loss to optimize the support- and query-level interaction relations between instances and strengthen label knowledge propagation to further improve performance.

- Extensive experiments show that our proposed method outperforms strong baselines and obtains significant performance on the few-shot MID task.

2 Related Work

2.1 Few-shot Learning

Few-shot learning has achieved great progress in Computer Vision (CV) [Ouyang *et al.*, 2022; Chen *et al.*, 2025] and Natural Language Processing (NLP) [Li *et al.*, 2023; Chan *et al.*, 2023; Zhou *et al.*, 2025]. Meta-learning is the mainstream method [Zhang *et al.*, 2022b; Liang *et al.*, 2023; Zhou *et al.*, 2024b; Zhou *et al.*, 2024a] for few-shot learning, and it includes model-based [Tsendsuren and Hong, 2017], optimization-based [Lee *et al.*, 2019], and metric-based [Wang *et al.*, 2021; Lv *et al.*, 2021; Zhao *et al.*, 2022a; Chen *et al.*, 2023] methods. Among these methods, metric-based methods are the most popular and potential research

works. However, few-shot learning methods mostly focus on single-label prediction [Zhang *et al.*, 2022a; Yehudai *et al.*, 2023; Du *et al.*, 2023], where they assign an intent to each utterance. Obviously, these methods become problematic in real-world scenarios because an utterance may have multiple intents. Few-shot MID could be regarded as a multi-label classification task. Previous few-shot multi-label classifications focus on CV domain [Yan *et al.*, 2022], audio domain [Cheng *et al.*, 2019], and sentiment analysis [Hu *et al.*, 2021; Zhao *et al.*, 2023a]. Few-shot MID is still in its infancy, and a few methods are available.

2.2 Few-shot Multi-label Intent Detection

Existing methods mostly focus on a two-stage pipeline. Hou et al. [2021] learn instance representations for multiple labels in the first stage. The second stage computes the relevance scores between the learned representations and different labels. Then, it uses a meta-calibrated dynamic threshold to perform multi-label classification based on relevance score ranking. However, they ignore the interaction relations between instances and do not handle the identical representation of instances with multiple labels. Graph Neural Networks (GNNs) [Zhao *et al.*, 2023b; Sun *et al.*, 2023; Chen *et al.*, 2022] have shown great progress in modeling the relationship between instances using graph structure. Zhang et al. [2023] propose a dual-class knowledge propagation network, which uses an instance-level GNN and a class-level GNN to learn instance representations. Then, they use a label count estimation to predict the label number of each instance. However, these methods rely on the representation classification in a pipeline, inevitably causing error propagation.

Different from the aforementioned methods, our proposed method is the first to propose multi-label joint learning for few-shot MID in an end-to-end manner. We design an instance relation learning network with label knowledge propagation to explicitly explore the interaction relations between instances, which eliminates the negative effect caused by error propagation and works well in low-resource domains.

3 Methodology

3.1 Task Formulation

We formulate the few-shot MID task in meta-learning, which is commonly employed in current studies [Yu *et al.*, 2022; Fang *et al.*, 2023; Ma *et al.*, 2023] to generalize prior knowledge to a low-resource target domain from resource-rich source domains. Our proposed method trains the model in a source domain $\mathcal{D}_{source}$ by building a collection of meta-tasks and tests it in a target domain $\mathcal{D}_{target}$ with other meta-tasks. Each meta-task consists of a support set and a query set. Specifically, the support set $\mathcal{S} = \{(x_i, y_i)\}_{i=1}^{N \times K}$ includes N classes (N-way), and each class has K instances (K-shot). As a multi-label task, few-shot MID could assign more than one label to each instance. Briefly, for an instance x_i, its corresponding label is $y_i = \{y_i^1, y_i^2, ..., y_i^N\} \in \{1, 0\}^N$, where $y_i^{c_k} = 1$ represents x_i belongs to class c_k. Besides, $\mathcal{Q} = \{(x_j)\}_{j=1}^T$ is called the query set, and T is the number of query instances. The training objective is to minimize the query loss of $\mathcal{Q}$ based on $\mathcal{S}$ in $\mathcal{D}_{source}$. In the testing phase,
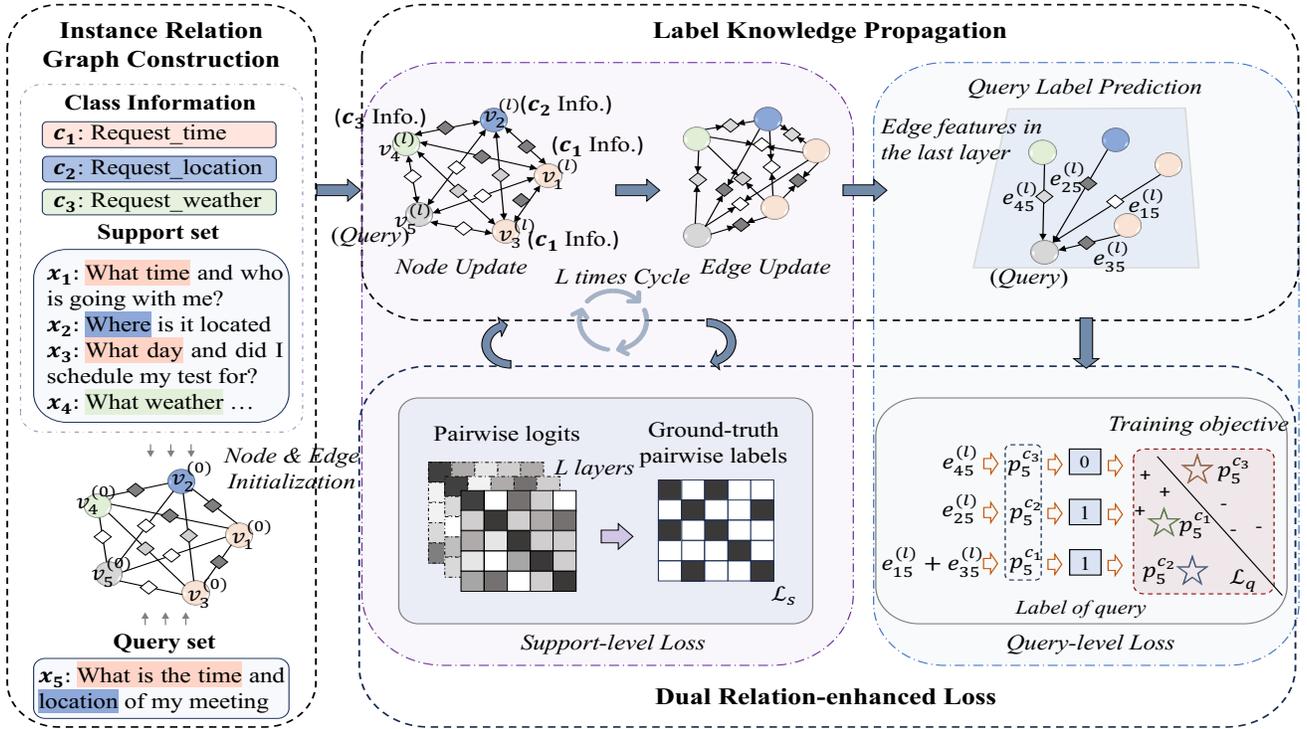
Figure 2: The overall architecture of the proposed method.

we construct $\mathcal{S}$ and $\mathcal{Q}$ in $\mathcal{D}_{target}$. Therefore, the proposed method predicts the query labels of $\mathcal{Q}$ based on $\mathcal{S}$ to verify the effectiveness of the model in the target domain $\mathcal{D}_{target}$.

3.2 Overall Framework

This section describes the proposed method for few-shot MID, as illustrated in Figure 2. We use the support set and the query set to construct a fully connected instance relation graph. Each node is an utterance, and each edge represents the relation strength between two connected nodes. In the graph network, our proposed method explicitly models the relationship between intra- and inter-class instances, which propagates label knowledge from support instances to query instances. Besides, a dual relation-enhanced loss is introduced to promote inference by enhancing support-level relation strength and query-level relation strength.

3.3 Instance Feature Extraction

Given an utterance x with n words, the utterance is defined as $x = \{w_1, w_2, ..., w_n\}$. We use a pre-trained text encoder like BERT [Devlin *et al.*, 2019] to encode each word and generate hidden states $H^x = [h_1, h_2, ..., h_n] \in \mathbb{R}^{n \times d}$. d is the dimension of hidden states. Besides, the corresponding class description c with m words is encoded into hidden states $H^c = [h_1, h_2, ..., h_m] \in \mathbb{R}^{m \times d}$.

Support Instance Feature

As an utterance has multiple intents (a.k.a. labels), it could be classified into different classes. Support utterances for each class could express better class-related intent semantics. To extract class-related semantic features, we concatenate a

support utterance and its class text $H^s = [H^x : H^c] \in \mathbb{R}^{(n+m) \times d}$ and utilize a self-attention mechanism [Yan *et al.*, 2020; Liu *et al.*, 2022; Du *et al.*, 2022] to learn attention weights A for each word, which generates the final representation of the support utterance. Specifically,

$$A = \text{softmax}(W_2 \tanh(W_1 H^s + b_1) + b_2), \quad (1)$$

$$u_s = W_3 \text{ReLU}(H^s A^T) + b_3, \quad (2)$$

where $u_s \in \mathbb{R}^d$ is the final feature representation of the support utterance. W_1, W_2, W_3, b_1, b_2 and b_3 are trainable parameters.

Query Instance Feature

For each query utterance, a mean pooling layer [Chen and Sun, 2023] is utilized to calculate the average of hidden states across all words, serving as the feature representation of the query utterance.

$$u_q = \text{MeanPooling}(H^x). \quad (3)$$

where $u_q \in \mathbb{R}^d$ is the query feature.

3.4 Instance Relation Graph Construction

Graph structure is beneficial for exploring the interaction relations between instances. Therefore, the proposed method utilizes the graph network to model the relationship between instances. As illustrated in Figure 2, the instance relation graph is denoted as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the node set and $\mathcal{E}$ is the edge set. The nodes in $\mathcal{V}$ correspond to instance features, and the edge $e_{ij} \in \mathcal{E}$ encodes the relationship between the ith instance and the jth instance.

Node Initialization

The nodes in the graph $\mathcal{G}$ are defined as $\mathcal{V} = \{v_i\}_{i=1}^M$, where $M = |\mathcal{S}| + |\mathcal{Q}|$, and v_i is the feature representation of the ith instance in the graph network. Nodes are initialized as $v_i^{(0)} = f_{emb}(x_i)$, where $f_{emb}(*)$ is the feature extraction function.

Edge Initialization

The edges in the graph $\mathcal{G}$ are defined as $\mathcal{E} = \{e_{ij}\}_{i,j}^M$ and represent the relationship between nodes. Besides, e_{ij} expresses the relation strength between two connected nodes and could be regarded as a probability that v_i and v_j are from the same class, reflecting intra-class similarity and inter-class dissimilarity. To better model the relationship between instances, the instance features are projected as key and query [VS *et al.*, 2023]. For the ith instance, its key (k_i), query (q_i) and node pairwise logits (e_{ij}) are as follows:

$$k_i^{(0)} = W_k^{(0)} \cdot v_i^{(0)}, \tag{4}$$

$$q_i^{(0)} = W_q^{(0)} \cdot v_i^{(0)}, \tag{5}$$

$$r_{ij}^{(0)} = q_i^{(0)}(k_j^{(0)})^T, \tag{6}$$

where $W_k^{(0)}$ and $W_q^{(0)}$ are trainable parameters. The edge features are written as follows:

$$e_{ij}^{(0)} = \begin{cases} r_{ij}^{(0)}, & if \ \{v_i^{(0)}, \ v_j^{(0)}\} \in \mathcal{S} \\ 0, & otherwise \end{cases} \tag{7}$$

where $e_{ij}^{(0)}$ represents the edge feature between nodes v_i and v_j in the 0th layer.

3.5 Label Knowledge Propagation

The proposed method incorporates class information into representations of support instances in Equation 1 and Equation 2. With message passing, the proposed method propagates label knowledge between support and query instances to enrich node and edge features.

Node Update

The proposed method propagates label knowledge from neighbors to enrich node features.

$$\tilde{v}_i^{(l)} = \sum_j \frac{e_{ij}^{(l-1)}}{\sum_k e_{ik}^{(l-1)}} v_j^{(l-1)}, \tag{8}$$

$$v_i^{(l)} = W_4 \text{ReLU}(\text{MLP}(\tilde{v}_i^{(l)}) + v_i^{(l-1)}) + b_4, \tag{9}$$

where $v_i^{(l)}$ is the ith node feature in the lth layer. MLP$(*)$ is a multi-layer perceptron. W_4 and b_4 are trainable parameters.

Edge Update

With the graph network, the proposed method propagates label knowledge to adjust edge features from the latest node features. The edge features are written as follows:

$$k_i^{(l)} = W_k^{(l)} \cdot v_i^{(l)}, \tag{10}$$

$$q_i^{(l)} = W_q^{(l)} \cdot v_i^{(l)}, \tag{11}$$

$$e_{ij}^{(l)} = q_i^{(l)}(k_j^{(l)})^T, \tag{12}$$

where $W_k^{(l)}$ and $W_q^{(l)}$ are trainable parameters.

3.6 Dual Relation-enhanced Loss

The dual relation-enhanced loss includes a support-level loss and a query-level loss to promote label knowledge propagation and multi-label inference.

Support-level Loss

The support-level loss enhances the relation strength between instances with the same label connected to the support instance and weakens the relation strength between those with different labels. The ground-truth label relation of an instance pair is given as follows.

$$y_{ij} = \begin{cases} 1, & if \ y_i = y_j, v_i \in \mathcal{S}, v_j \in \{\mathcal{S}, \mathcal{Q}\}, \\ 0, & if \ y_i \neq y_j, v_i \in \mathcal{S}, v_j \in \{\mathcal{S}, \mathcal{Q}\}, \end{cases} \tag{13}$$

where y_i and y_j are the labels of v_i and v_j, respectively. $y_{ij} = 1$ denotes that v_i and v_j belong to the same class.

Any i^{th} support instance is considered an anchor, where $v_i \in \mathcal{S}$. The remaining instances are constructed a set as $J = \{\mathcal{S}, \mathcal{Q}\} \backslash \{v_i\}$. A positive set is $\Omega_{pos}^s = \{e_{ij}^{(l)} | y_{ij} = 1, i \neq j\}$, and a negative set is $\Omega_{neg}^s = \{e_{ij}^{(l)} | y_{ij} = 0, i \neq j\}$. The support-level loss could be formulated as follows.

$$\mathcal{L}_s = \sum_{l=0} \frac{1}{|\mathcal{S}|} \sum_{v_i \in \mathcal{S}} (\log(1 + \sum_{e_{ij}^{(l)} \in \Omega_{neg}^s} \exp(e_{ij}^{(l)})) +$$
$$\log(1 + \sum_{e_{ij}^{(l)} \in \Omega_{pos}^s} \exp(-e_{ij}^{(l)}))), \tag{14}$$

From the above formula, we can observe that the relation strength scores of connected instances derived from the same class are greater than value 0, while those with different classes are less than value 0. In this way, the support-level loss encourages the target node to learn information from similar neighbors and pushes it away from dissimilar neighbors to guide label knowledge propagation effectively.

Query-level Loss

The edge feature e_{ij} is regarded as a probability that two connected nodes v_i and v_j are from the same class. After L layers of network propagation, the latest edge features from support nodes to query nodes are used for query inference. Specifically, each query node could be classified by final edge voting with support labels. The prediction probability that a query node v_i belongs to class c_k could be formulated as $p_i^{c_k}$.

$$p_i^{c_k} = \sum_{v_j \in S} e_{ji}^{(l)} \delta(y_j^s = c_k), \tag{15}$$

where y_j^s is the label of the support node v_j. e_{ji} is the edge between the support node v_j and the query node v_i. $\delta(y_j^s = c_k)$ is the Kronecker delta function that is equal to one when $y_j^s = c_k$ and zero otherwise.

Considering that a query may contain multiple user intents, the positive set for the ith query node is defined as $\Omega_{pos}^q = \{p_i^{c_k} | y_i^q = c_k\}$, which indicates the prediction probability set

between a query node and its corresponding classes. In contrast, a negative set is defined as $\Omega_{neg}^q = \{p_i^{c_k} | y_i^q \neq c_k\}$, which indicates the prediction probability set between a query node and its irrelevant classes. The training objective is used to minimize the following loss function.

$$
\mathcal{L}_q = \frac{1}{|\mathcal{Q}|} \sum_{i=1}^{|\mathcal{Q}|} (\log(1 + \sum_{p_i^{c_k} \in \Omega_{neg}^q} \exp(p_i^{c_k})) + \\
\log(1 + \sum_{p_i^{c_k} \in \Omega_{pos}^q} \exp(-p_i^{c_k}))),
\tag{16}
$$

The aforementioned loss function aims to ensure that the scores in Ω_{pos}^q are greater than value 0 and the scores in Ω_{neg}^q are less than value 0. Therefore, the label $\hat{y}_i$ of the query instance is formulated as follows:

$$
\hat{y_i}^{c_k} = \begin{cases} 1, & if \ p_i^{c_k} > 0 \\ 0, & if \ p_i^{c_k} < 0 \end{cases}
\tag{17}
$$

where $\hat{y_i}^{c_k} = 1$ shows that v_i is predicted to class c_k, whereas $\hat{y_i}^{c_k} = 0$ shows that v_i does not belong to c_k.

3.7 Training Objective

So far, we have introduced an instance relation learning network to effectively tackle the few-shot MID task. The overall training objective is written as follows:

$$
\mathcal{L} = \alpha \mathcal{L}_s + \beta \mathcal{L}_q,
\tag{18}
$$

where α and β are hyper-parameters.

4 Experiments

4.1 Experimental Setup

Datasets. We conduct experiments on the benchmark dataset TourSG [Williams *et al.*, 2012]. TourSG comprises six domains: Itinerary (It), Accommodation (Ac), Attraction (At), Food (Fo), Transportation (Tr), and Shopping (Sh). The detailed statistics are presented in Table 1.

Evaluation Metric. Following previous works [Hou *et al.*, 2021; Zhang *et al.*, 2023], we use AUC and Macro-F1 scores to evaluate the performance of our proposed method.

Implementation Details. The proposed method is implemented with PyTorch (version 1.10.0) on a single GPU (RTX 3090 Ti) with CUDA version 11.3. We use the pre-trained language model BERT-base [Devlin *et al.*, 2019] as our encoder for H (see Feature Extraction). The AdamW optimizer trains the model with a learning rate of 5e-5. Meanwhile, we use the GradualWarmupScheduler to optimize the learning rate and set the warmup proportion to 0.05. The hyper-parameters α and β are fixed as 0.1 and 1. We set L to 2 as the number of cycles for the updates in the overall architecture. We test and validate the model on two domains, respectively, and train it on the remaining domains. 100 meta-tasks are randomly sampled for training, validation, and testing in every epoch. We evaluate Macro-F1 scores in each meta-task and obtain the average F1 scores for all meta-tasks. Finally, the results of the test domain are reported when the validation domain obtains the best results.

Dataset	It	Ac	At	Fo	Tr	Sh
#cla.	15	17	18	18	17	16
#ins.	397	1839	6162	2154	2493	1278

Table 1: Statistics of datasets. #cla. denotes the number of intents, and #ins. indicates the number of utterances.

4.2 Baselines

We compare the proposed method with a series of strong baselines. AWATT [Hu *et al.*, 2021] designs a support-set attention mechanism and a query-set attention mechanism for learning representations of multi-label instances. Then, they use reinforcement learning to obtain a fixed threshold to identify multiple categories from these representations. HATT [Gao *et al.*, 2019] and LDF [Zhao *et al.*, 2022b] use an empirical threshold to achieve multi-label results. Besides, HATT utilizes instance-level and feature-level attention mechanisms to learn robust instance representations. LDF uses contrastive learning to learn different representations of multi-label instances. CTLR [Hou *et al.*, 2021] combines an instance and its label text to learn representations for multiple labels. Then, it leverages a meta-calibrated method to learn a dynamic threshold. In addition, LPN [Liu *et al.*, 2022] and DCKPN [Zhang *et al.*, 2023] use label count estimation. They use an adaptive neural network to predict the label number of each instance. Besides, we also compare our method with Large Language Models (LLMs), such as Llama3-70b [Touvron *et al.*, 2023] and GhatGLM4 [GLM *et al.*, 2024].

4.3 Main Results

We conduct extensive experiments and report results in Tables 2, 3, 4 and 5. The best results are highlighted in bold, with the following observations.

(1) Our proposed method achieves an average of 5.24% AUC and 7.62% Macro-F1 improvement in the 5-way scenario. In Table 2 and Table 3, the proposed method improves upon the most competitive baseline DCKPN by 6.49%-9.88% AUC and 10.07%-17.21% Macro-F1 scores in the 5-way 1-shot setting, with an average improvement of 7.60% AUC and 11.89% Macro-F1 scores. In terms of the 5-way 3-shot setting, the proposed method improves upon the most competitive baseline DCKPN by 1.57%-5.79% AUC and 0.77%-8.47% Macro-F1 scores, with an average improvement of 2.89% AUC and 3.36% Macro-F1 scores.

(2) Furthermore, in Table 4 and Table 5, our proposed method achieves an average of 9.60% AUC and 5.73% Macro-F1 scores improvement in the 10-way scenario. In terms of the 10-way 1-shot setting, the proposed method improves upon the most competitive baseline DCKPN by 10.34%-12.69% AUC and 4.97%-15.68% Macro-F1 scores, with an average improvement of 11.50% AUC and 10.49% Macro-F1 scores. In terms of the 10-way 3-shot setting, the proposed method improves upon the most competitive baseline DCKPN by 6.46%-8.87% AUC, with an average improvement of 7.71% AUC. Compared to other methods, our 1-shot setting outperforms the 3-shot setting. Our proposed method focuses on edges in the graph network, using these

Models	It		Ac		At		Fo		Sh		Tr	
	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot
Llama3	-	-	-	-	-	-	-	-	-	-	-	-
GhatGLM4	-	-	-	-	-	-	-	-	-	-	-	-
CTLR	63.32	69.32	60.37	71.24	64.87	71.31	62.92	69.65	63.90	69.56	62.83	70.70
HATT	68.96	73.03	79.93	80.97	78.53	83.35	79.35	84.35	76.74	80.90	81.02	84.78
LPN	65.89	73.51	73.36	79.24	74.71	82.00	69.04	79.27	70.54	78.58	70.58	81.17
AWATT	68.76	76.99	80.26	84.81	78.18	85.44	79.13	86.50	77.72	84.38	80.33	87.06
LDF	68.46	75.88	80.46	84.17	80.38	85.18	79.50	86.47	77.08	84.49	79.83	87.83
DCKPN	73.21	76.80	81.23	83.83	79.86	85.00	81.55	85.86	78.46	83.32	82.37	86.66
Ours	**83.09**	**82.59**	**88.26**	**86.43**	**87.08**	**86.57**	**88.04**	**87.76**	**86.89**	**86.28**	**88.89**	**89.20**

Table 2: Comparison of AUC in 5-way scenarios.

Models	It		Ac		At		Fo		Sh		Tr	
	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot
Llama3	48.91	48.45	54.23	49.57	53.07	54.35	57.54	52.56	53.66	51.25	52.35	51.77
GhatGLM4	32.40	33.56	39.26	40.51	38.02	47.67	46.44	41.63	35.40	31.01	36.49	37.18
CTLR	39.31	40.30	38.00	40.77	37.11	39.59	37.88	40.14	38.03	40.22	38.44	40.37
HATT	39.27	46.01	50.47	55.40	28.28	58.03	39.33	60.69	43.92	56.05	43.50	62.89
LPN	31.98	42.91	39.69	49.90	40.94	53.45	35.00	50.13	37.50	49.65	36.61	53.28
AWATT	40.25	50.61	53.52	62.22	50.88	62.49	52.24	64.62	50.61	61.95	54.90	66.47
LDF	39.83	49.64	54.10	61.07	53.89	62.18	52.52	65.01	49.56	62.06	54.06	67.55
DCKPN	45.73	53.35	57.89	62.37	55.13	64.52	58.39	66.74	53.87	62.48	59.95	68.73
Ours	**62.94**	**61.82**	**69.15**	**65.87**	**65.20**	**65.29**	**68.78**	**68.24**	**64.98**	**65.31**	**71.24**	**71.80**

Table 3: Comparison of Macro-F1 scores in 5-way scenarios.

Models	It		Ac		At		Fo		Sh		Tr	
	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot
Llama3	-	-	-	-	-	-	-	-	-	-	-	-
GhatGLM4	-	-	-	-	-	-	-	-	-	-	-	-
CTLR	62.01	68.76	65.93	71.69	65.32	73.43	62.73	69.87	62.66	69.92	63.67	70.64
HATT	67.05	72.44	77.41	78.84	76.83	81.42	76.99	81.60	74.57	79.44	77.02	82.39
LPN	65.40	74.09	72.90	79.18	73.70	81.59	68.94	79.50	69.17	78.21	69.52	80.55
AWATT	66.27	73.98	77.64	81.70	77.79	83.68	77.23	84.16	74.71	81.84	77.48	84.72
LDF	66.58	76.14	77.86	81.46	77.79	83.96	77.39	84.00	75.23	81.53	77.14	84.83
DCKPN	69.50	71.88	76.03	77.62	75.20	78.72	76.84	79.28	72.64	77.85	75.92	81.22
Ours	**81.39**	**80.75**	**87.20**	**86.01**	**86.89**	**85.99**	**87.18**	**86.73**	**85.33**	**85.64**	**87.14**	**87.68**

Table 4: Comparison of AUC in 10-way scenarios.

edges to directly indicate whether two instances belong to the same label. Therefore, each query node is classified based on final edge voting with support labels. In 1-shot scenarios, fewer edges make optimization easier within the dual relation enhancement loss compared to in 3-shot scenarios, which is why 1-shot scenarios outperform 3-shot scenarios. This advantage demonstrates that our proposed method can achieve better results with fewer samples. However, DCKPN or LPN focuses on representation classification and ignores explicit interaction between instances, leading to low performance in 1-shot scenarios due to error propagation. Therefore, our proposed method guides the multi-label inference effectively.

(3) Compared to LLMs, our proposed method consistently outperforms GhatGLM4 and Llama3. For few-shot MID, the task requires LLMs to extract all relevant intents from a given utterance. However, LLMs cannot understand dialogue diversity in different domains. Even with a few demonstration examples, LLMs may generate outputs that are not fully aligned with the expected labels. In the example "*What time is my meeting and what day?*", LLMs might only generate outputs like "*request time*", whereas the ground-truth labels are "*request time*" and "*request day*". These incomplete outputs can degrade the model's performance, making it difficult for LLMs to handle multi-label tasks effectively.

Models	It		Ac		At		Fo		Sh		Tr	
	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot	1-shot	3-shot
Llama3	41.90	38.42	48.39	45.42	48.33	46.22	50.84	44.07	44.64	34.58	42.36	31.81
GhatGLM4	26.62	29.66	34.95	38.00	35.60	37.53	31.97	37.40	30.25	28.36	32.11	25.79
CTLR	26.46	28.72	26.06	28.41	24.56	28.13	24.72	27.27	25.45	27.34	25.96	27.91
HATT	21.58	36.39	30.25	45.26	36.74	47.34	25.30	48.08	29.81	44.55	32.66	50.52
LPN	22.46	32.15	27.44	39.19	29.12	43.41	33.33	38.73	24.81	38.05	33.67	42.50
AWATT	28.93	37.20	41.85	49.51	41.01	51.32	40.99	52.25	39.16	48.81	43.62	55.17
LDF	29.40	40.13	42.24	**49.57**	40.89	**51.60**	41.03	52.02	39.00	**48.90**	43.07	54.73
DCKPN	34.20	39.18	41.93	48.00	39.29	48.83	46.58	**53.31**	38.48	48.57	39.66	55.00
Ours	**47.45**	**47.76**	**51.34**	47.48	**49.00**	47.10	**51.55**	52.99	**48.43**	47.74	**55.34**	**55.65**

Table 5: Comparison of Macro-F1 scores in 10-way scenarios.

Model	AUC	Δ AUC	F1	Δ F1
Full model	**88.04**		**68.78**	
w/o label knowledge	76.50	-11.54	53.42	-15.36
w/o node pairwise logits	86.39	-1.65	66.84	-1.94
w/o dual loss	78.62	-9.42	44.75	-24.03
w/o support-level loss	85.87	-2.17	66.57	-2.21

Table 6: An ablation study on the 5-way 1-shot scenario.

4.4 Impact of Shot Number

To investigate the impact of the shot number for each class, we evaluate our proposed method with one to eleven shots in the food domain. Figure 3 shows the performances of methods in 5-way scenarios. The performances of DCKPN, LDF, AWATT, and HATT improve at first then fluctuate. The reason is that they focus on representation classification and rely on the number of labeled data to learn great instance representations. However, these methods underperform in the 1-shot setting. Our proposed method focuses on edges in the graph structure and utilizes the relation strength between instances to guide multi-label inference. As discussed in the main results, optimizing the training objectives in fewer instances would be easy, which achieves great performance. Our proposed method obtains the best results on the 1-shot setting, demonstrating the effectiveness of the model.

4.5 Ablation Study

In Table 6, we conduct an ablation study. (1) "w/o label knowledge". We remove class descriptions in the support set. The negative results suggest that class descriptions promote label knowledge propagation in the network. (2) "w/o node pairwise logits". We replace node pairwise logits with cosine similarity as edge features in the network. The low performance indicates that node pairwise logits play a positive effect by utilizing key and query of instances. (3) "w/o dual loss". After we remove the dual relation-enhanced loss, the performance drops a lot. The negative results suggest that enhancing support- and query-level relation strength is a key point for multi-label prediction. (4) "w/o support-level loss". Though we only remove the support-level loss, the performance is reduced. In conclusion, the complete model outperforms all ablation studies and achieves the best performance.

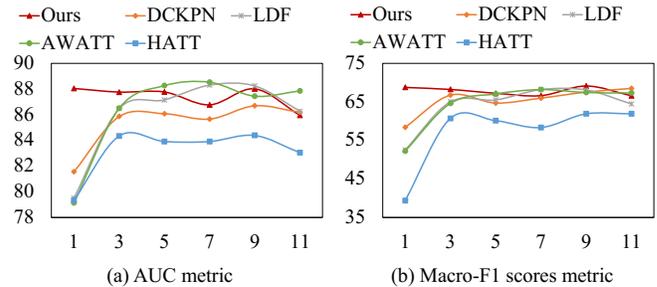(a) AUC metric (b) Macro-F1 scores metric

Figure 3: The impact of shot number.

5 Conclusion

We propose a multi-label joint learning method for few-shot MID, addressing the negative effect caused by error propagation. The proposed method explicitly models the relationship between instances and utilizes the relation strength between an instance pair to directly indicate whether these two instances belong to the same label, guiding multi-label inference effectively in an end-to-end manner. Besides, we design a dual relation-enhanced loss, which enhances interaction relations between support- and query-level instances and strengthens label knowledge propagation to further improve performance. Experiments show that we significantly outperform strong baselines, esp. the 1-shot setting.

6 Limitation

We incorporate class descriptions into the graph network to facilitate label knowledge propagation. Given that class descriptions are often brief and may provide limited context, their contribution to the overall model may be constrained. To improve performance, we plan to integrate more comprehensive external knowledge sources (e.g., domain-specific ontologies) into the graph construction, ultimately leading to better classification and prediction outcomes. Besides, our proposed method focuses on the few-shot MID task in dialogue systems. In the following work, we will explore other applications, such as health care.

Acknowledgements

References

[Chan *et al.*, 2023] Jun Shern Chan, Michael Pieler, Jonathan Jao, Jérémy Scheurer, and Ethan Perez. Few-shot adaptation works with UnpredicTable data. In *Proc. of ACL*, pages 1806–1842, 2023.

[Chen and Sun, 2023] Zhaozheng Chen and Qianru Sun. Extracting class activation maps from non-discriminative features as well. In *Proc. of CVPR*, pages 3135–3144, 2023.

[Chen *et al.*, 2022] Junru Chen, Yang Yang, Tao Yu, Yingying Fan, Xiaolong Mo, and Carl Yang. Brainnet: Epileptic wave detection from seeg with hierarchical graph diffusion learning. In *Proc. of ACM SIGKDD*, pages 2741–2751, 2022.

[Chen *et al.*, 2023] Xiudi Chen, Hui Wu, and Xiaodong Shi. Consistent prototype learning for few-shot continual relation extraction. In *Proc. of ACL*, pages 7409–7422, 2023.

[Chen *et al.*, 2025] Wei Chen, Shigui Li, Jiacheng Li, Junmei Yang, John Paisley, and Delu Zeng. Dequantified diffusion schrödinger bridge for density ratio estimation. *arXiv preprint arXiv:2505.05034*, 2025.

[Cheng *et al.*, 2019] Kai-Hsiang Cheng, Szu-Yu Chou, and Yi-Hsuan Yang. Multi-label few-shot learning for sound event recognition. In *Proc. of MMSP*, pages 1–5, 2019.

[Devlin *et al.*, 2019] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proc. of ACL*, 2019.

[Du *et al.*, 2022] Shian Du, Yihong Luo, Wei Chen, Jian Xu, and Delu Zeng. To-flow: Efficient continuous normalizing flows with temporal optimization adjoint with moving speed. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12570–12580, 2022.

[Du *et al.*, 2023] Jiangshu Du, Congying Xia, Wenpeng Yin, Tingting Liang, and Philip S Yu. All labels together: Low-shot intent detection with an efficient label semantic encoding paradigm. In *Proc. of ACL*, 2023.

[Fang *et al.*, 2023] Jinyuan Fang, Xiaobin Wang, Zaiqiao Meng, Pengjun Xie, Fei Huang, and Yong Jiang. MANNER: A variational memory-augmented model for cross domain few-shot named entity recognition. In *Proc. of ACL*, pages 4261–4276, 2023.

[Gao *et al.*, 2019] Tianyu Gao, Xu Han, Zhiyuan Liu, and Maosong Sun. Hybrid attention-based prototypical networks for noisy few-shot relation classification. In *Proc. of AAAI*, volume 33, pages 6407–6414, 2019.

[GLM *et al.*, 2024] Team GLM, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Diego Rojas, Guanyu Feng, Hanlin Zhao, Hanyu Lai, Hao Yu, Hongning Wang, Jiadai Sun, Jiajie Zhang, Jiale Cheng, Jiayi Gui, Jie Tang, Jing Zhang, Juanzi Li, Lei Zhao, Lindong Wu, Lucen Zhong, Mingdao Liu, Minlie Huang, Peng Zhang, Qinkai Zheng, Rui Lu, Shuaiqi Duan, Shudan Zhang, Shulin Cao, Shuxun Yang, Weng Lam Tam, Wenyi Zhao, Xiao Liu, Xiao Xia, Xiaohan Zhang, Xiaotao Gu, Xin Lv, Xinghan Liu, Xinyi Liu, Xinyue Yang, Xixuan Song, Xunkai Zhang, Yifan An, Yifan Xu, Yilin Niu, Yuantao Yang, Yueyan Li, Yushi Bai, Yuxiao Dong, Zehan Qi, Zhaoyu Wang, Zhen Yang, Zhengxiao Du, Zhenyu Hou, and Zihan Wang. Chatglm: A family of large language models from glm-130b to glm-4 all tools, 2024.

[Han *et al.*, 2021] Janghoon Han, Taesuk Hong, Byoungjae Kim, Youngjoong Ko, and Jungyun Seo. Fine-grained post-training for improving retrieval-based dialogue systems. In *Proc. of ACL*, pages 1549–1558, 2021.

[Hou *et al.*, 2021] Yutai Hou, Yongkui Lai, Yushan Wu, Wanxiang Che, and Ting Liu. Few-shot learning for multi-label intent detection. In *Proc. of AAAI*, pages 13036–13044, 2021.

[Hu *et al.*, 2021] Mengting Hu, Shiwan Zhao, Honglei Guo, Chao Xue, Hang Gao, Tiegang Gao, Renhong Cheng, and Zhong Su. Multi-label few-shot learning for aspect category detection. In *Proc. of ACL*, 2021.

[Lee *et al.*, 2019] Kwonjoon Lee, Subhransu Maji, Avinash Ravichandran, and Stefano Soatto. Meta-learning with differentiable convex optimization. In *Proc. of CVPR*, pages 10657–10665, 2019.

[Li *et al.*, 2023] Peng Li, Tianxiang Sun, Qiong Tang, Hang Yan, Yuanbin Wu, Xuanjing Huang, and Xipeng Qiu. CodeIE: Large code generation models are better few-shot information extractors. In *Proc. of ACL*, pages 15339–15353, 2023.

[Liang *et al.*, 2023] Bin Liang, Xiang Li, Lin Gui, Yonghao Fu, Yulan He, Min Yang, and Ruifeng Xu. Few-shot aspect category sentiment analysis via meta-learning. *ACM Transactions on Information Systems*, 41(1):1–31, 2023.

[Liu *et al.*, 2022] Han Liu, Feng Zhang, Xiaotong Zhang, Siyang Zhao, Junjie Sun, Hong Yu, and Xianchao Zhang. Label-enhanced prototypical network with contrastive learning for multi-label few-shot aspect category detection. In *Proc. of ACM SIGKDD*, pages 1079–1087, 2022.

[Lv *et al.*, 2021] Hui Lv, Chen Chen, Zhen Cui, Chunyan Xu, Yong Li, and Jian Yang. Learning normal dynamics in videos with meta prototype network. In *Proc. of CVPR*, pages 15425–15434, 2021.

[Ma *et al.*, 2023] Ruotian Ma, Zhang Lin, Xuanting Chen, Xin Zhou, Junzhe Wang, Tao Gui, Qi Zhang, Xiang Gao, and Yunwen Chen. Coarse-to-fine few-shot learning for named entity recognition. In *Proc. of ACL*, pages 4115–4129, 2023.

[Ouyang *et al.*, 2022] Cheng Ouyang, Carlo Biffi, Chen Chen, Turkay Kart, Huaqi Qiu, and Daniel Rueckert. Self-supervised learning for few-shot medical image segmentation. *IEEE Transactions on Medical Imaging*, 41(7):1837–1848, 2022.

[Pham *et al.*, 2023] Thinh Pham, Chi Tran, and Dat Quoc Nguyen. Misca: A joint model for multiple intent detection and slot filling with intent-slot co-attention. In *Proc. of EMNLP*, 2023.

[Sun *et al.*, 2023] Xiangguo Sun, Hong Cheng, Bo Liu, Jia Li, Hongyang Chen, Guandong Xu, and Hongzhi Yin. Self-supervised hypergraph representation learning for sociological analysis. *IEEE Transactions on Knowledge and Data Engineering*, 2023.

[Touvron *et al.*, 2023] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv:2302.13971*, 2023.

[Tsendsuren and Hong, 2017] Munkhdalai Tsendsuren and Yu Hong. Meta networks. In *International Conference on Machine Learning*, pages 2554–2563, 2017.

[Tu *et al.*, 2023] Nguyen Anh Tu, Hoang Thi Thu Uyen, Tu Minh Phuong, and Ngo Xuan Bach. Joint multiple intent detection and slot filling with supervised contrastive learning and self-distillation. In *Proc. of European Conference on Artificial Intelligence*, 2023.

[VS *et al.*, 2023] Vibashan VS, Poojan Oza, and Vishal M. Patel. Instance relation graph guided source-free domain adaptive object detection. In *Proc. of CVPR*, pages 3520–3530, 2023.

[Wang *et al.*, 2021] Runzhong Wang, Junchi Yan, and Xiaokang Yang. 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, 2021.

[Williams *et al.*, 2012] Jason D Williams, Antoine Raux, Deepak Ramachandran, and Alan Black. Dialog state tracking challenge handbook. *Silicon Valley, Microsoft Research*, 2012.

[Yan *et al.*, 2020] Guangfeng Yan, Fan Lu, Qimai Li, and Han Liu. Unknown intent detection using gaussian mixture model with an application to zero-shot intent classification. In *Proc. of ACL*, pages 1050–1060, 2020.

[Yan *et al.*, 2022] Kun Yan, Chenbin Zhang, Jun Hou, Ping Wang, Zied Bouraoui, Shoaib Jameel, and Steven Schockaert. Inferring prototypes for multi-label few-shot image classification with word vector guided attention. In *Proc. of AAAI*, pages 2991–2999, 2022.

[Yehudai *et al.*, 2023] Asaf Yehudai, Matan Vetzler, Yosi Mass, Koren Lazar, Doron Cohen, and Boaz Carmeli. Qaid: Question answering inspired few-shot intent detection. In *Proceedings of the International Conference on Learning Representations*, 2023.

[Yu *et al.*, 2022] Tianyuan Yu, Sen He, Yi-Zhe Song, and Tao Xiang. Hybrid graph neural networks for few-shot learning. In *Proc. of AAAI*, volume 36, pages 3179–3187, 2022.

[Zhang *et al.*, 2020] Yichi Zhang, Zhijian Ou, and Zhou Yu. Task-oriented dialog systems that consider multiple appropriate responses under the same context. In *Proc. of AAAI*, volume 34, pages 9604–9611, 2020.

[Zhang *et al.*, 2021] Jianguo Zhang, Trung Bui, Seunghyun Yoon, Xiang Chen, Zhiwei Liu, Congying Xia, Quan Hung Tran, Walter Chang, and Philip Yu. Few-shot intent detection via contrastive pre-training and fine-tuning. In *Proc. of EMNLP*, 2021.

[Zhang *et al.*, 2022a] Haode Zhang, Haowen Liang, Yuwei Zhang, Liming Zhan, Xiao-Ming Wu, Xiaolei Lu, and Albert Lam. Fine-tuning pre-trained language models for few-shot intent detection: Supervised pre-training and isotropization. In *Proc. of ACL*, 2022.

[Zhang *et al.*, 2022b] Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Cheekong Lee. Protgnn: Towards self-explaining graph neural networks. In *Proc. of AAAI*, volume 36, pages 9127–9135, 2022.

[Zhang *et al.*, 2023] Feng Zhang, Wei Chen, Fei Ding, and Tengjiao Wang. Dual class knowledge propagation network for multi-label few-shot intent detection. In *Proc. of ACL*, pages 8605–8618, 2023.

[Zhao *et al.*, 2022a] Fei Zhao, Yuchen Shen, Zhen Wu, and Xinyu Dai. Label-driven denoising framework for multi-label few-shot aspect category detection. *Proc. of EMNLP*, 2022.

[Zhao *et al.*, 2022b] Fei Zhao, Yuchen Shen, Zhen Wu, and Xinyu Dai. Label-driven denoising framework for multi-label few-shot aspect category detection. In *Proc. of EMNLP*, pages 2390–2402, Abu Dhabi, United Arab Emirates, December 2022. Proc. of ACL.

[Zhao *et al.*, 2023a] Shiman Zhao, Wei Chen, and Tengjiao Wang. Learning few-shot sample-set operations for noisy multi-label aspect category detection. In *Proc. of IJCAI*, 2023.

[Zhao *et al.*, 2023b] Shiman Zhao, Yutao Xie, Wei Chen, Tengjiao Wang, Jiahui Yao, and Jiabin Zheng. Metric-free learning network with dual relations propagation for few-shot aspect category sentiment analysis. *Proc. of ACL*, 2023.

[Zhou *et al.*, 2024a] Xiaoling Zhou, Ou Wu, and Nan Yang. Adversarial training with anti-adversaries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[Zhou *et al.*, 2024b] Xiaoling Zhou, Wei Ye, Zhemg Lee, Rui Xie, and Shikun Zhang. Boosting model resilience via implicit adversarial data augmentation. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 5653–5661, 2024.

[Zhou *et al.*, 2025] Xiaoling Zhou, Wei Ye, Zhemg Lee, Lei Zou, and Shikun Zhang. Valuing training data via causal inference for in-context learning. *IEEE Transactions on Knowledge and Data Engineering*, 2025.