Expert Systems

WILEY

Expert Systems

ORIGINAL ARTICLE

Few-Shot Contrastive Learning-Based Multi-Round Dialogue Intent Classification Method

Feng Wei[1] | Xu Zhang[2,3]

[1]China Telecom Corporation Limited Chongqing Branch, Chongqing, China | [2]Key Laboratory of Big Data Intelligent Computing, Chongqing University of Posts and Telecommunications, Chongqing, China | [3]Key Laboratory of Tourism Multisource Data Perception and Decision, Ministry of Culture and Tourism, Chongqing, China

Correspondence: Xu Zhang (zhangx@cqupt.edu.cn)

Received: 19 June 2024 | **Revised:** 7 October 2024 | **Accepted:** 14 October 2024

Funding: This work was supported by the Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant No. KJZD-M202400603), the Project of Key Laboratory of Tourism Multisource Data Perception and Decision, Ministry of Culture and Tourism, China (No. H2023009), the Key Cooperation Project of Chongqing Municipal Education Commission (No. HZ2021008).

Keywords: AI in convergence ICT | few-shot contrastive learning | intent classification | multi-round dialogue

ABSTRACT

Traditional text classification models face challenges in handling long texts and understanding topic transitions in dialogue scenarios, leading to suboptimal performance in automatic speech recognition (ASR)-based multi-round dialogue intent classification. In this article, we propose a few-shot contrastive learning-based multi-round dialogue intent classification method. First, the ASR texts are partitioned, and role-based features are extracted using a Transformer encoder. Second, refined sample pairs are forward-propagated, adversarial samples are generated by perturbing word embedding matrices and contrastive loss is applied to positive sample pairs. Then, positive sample pairs are input into a multi-round reasoning module to learn semantic clues from the entire scenario through multiple dialogues, obtain reasoning features, input them into a classifier to obtain classification results, and calculate multi-task loss. Finally, a prototype update module (PUM) is introduced to rectify the biased prototypes by using gated recurrent unit (GRU) to update the prototypes stored in the memory bank and few-shot learning (FSL) task. Experimental evaluations demonstrate that the proposed method outperforms state-of-the-art methods on two public datasets (DailyDialog and CM) and a private real-world dataset.

1 | Introduction

The concept of a Dialogue Act (DA), which is originated from 'illocutionary act' theory, is considered as a method of defining the semantic content and communicative function of a single utterance of dialogue (Duran, Battle, and Smith 2023). Dialogue intent classification (IC or DIC) is a pivotal component of task-oriented dialogue systems (Liu, Zhang, et al. 2021; Xu et al. 2023; Louvan and Magnini 2020; Firdaus et al. 2021), enabling machines to understand and respond to user intentions effectively. Conversational AI (CoAI) (Tian et al. 2022) bots for dialogue tasks have gradually been widely used in the past decade and gained much attention recently. As shown in Figure 1, a typical multi-round dialogue system should include many essential components for dialogue understanding, including automatic speech recognition (ASR) (He and Young 2003; Hayat et al. 2020), natural language understanding (NLU) (Liu, Eshghi, et al. 2021; Xiang et al. 2024), dialogue state tracking (DST) (Madotto et al. 2020), intent classification (IC) (Li 2022; Yuan 2023; Pang et al. 2022; Yan et al. 2024; Zhao et al. 2023; Sauer, Asaadi, and Küch 2022; Hou et al. 2021) and slot filling (SF) (Wu, Hovakimyan, and Hobbs 2023). In this article, we focus on the contents marked in blue, as shown in Figure 1. Traditional classification models are ineffective in dealing with

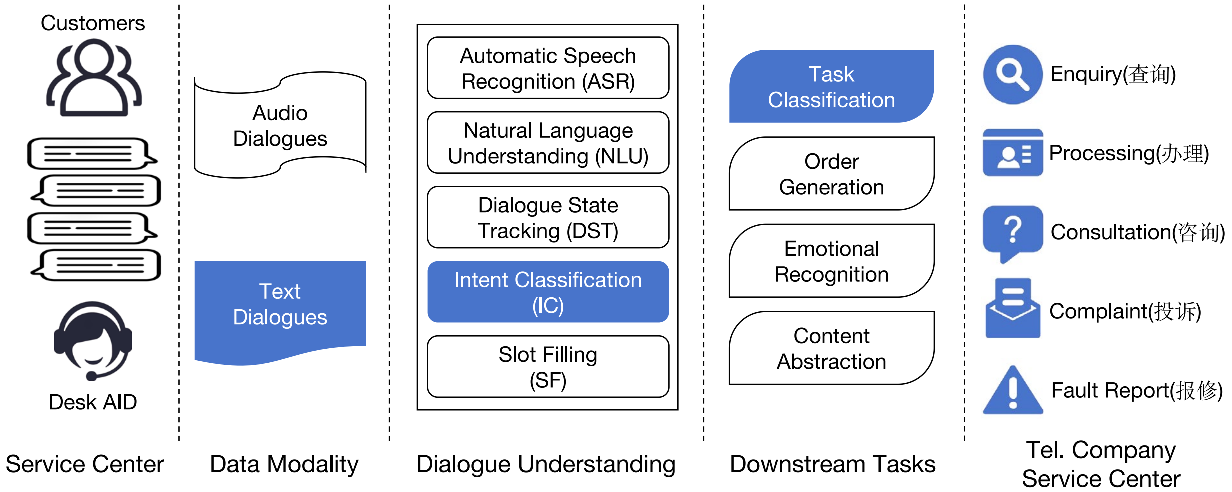

FIGURE 1 | A typical multi-round dialogue system for online/call service scenario.

these challenges. Recent advances in deep learning and neural network theory have shown promotion in text classification, however, multi-round dialogue intent classification in voice customer service scenarios is more challenging due to word repetitions, unclear intentions and ultra-long text content when the text is generated via automatic speech recognition (ASR).

Existing DIC methods often require large annotated datasets to obtain reasonable performance and avoid overfitting, which is not always available in practical scenarios. To mitigate this limitation, few-shot learning (FSL), which imitates the human's ability to learn from a few examples and adapt quickly to new tasks, has been proposed to train networks to understand the utterance of the user from only a few annotated data. Contrastive learning (Zhang et al. 2024; He et al. 2023; Xie et al. 2024), an effective self-supervised deep learning paradigm, learns activity representation by contrasting sample views created by varying data with augmentation methods. Most existing studies on dialogue classification focus on sentence-level or utterance-level intent recognition of user statements (Yan et al. 2024). Customer service dialogues belong to task-oriented multi-round dialogues aimed at solving specific problems in particular scenarios. As shown in Figure 1, a typical scenario in telecommunication company (e.g., China Telecom, 10,000) aims to classify user intent with a multi-round text dialogue into 5 business types, including *enquiry*, *processing*, *consultation*, *complaint* and *fault report*.

In this article, we propose a few-shot contrastive learning-based multi-round dialogue intent classification method. First, we partition the ASR text dialogues into long and short texts, extract role-based features, and encode them with a Transformer encoder. Then, the pre-processed sample pairs are processed with a forward-propagation, and perturbs word embedding matrices to generate adversarial samples, applies contrastive loss to positive sample pairs. We use knowledge distillation to train a feature extractor to alleviate the negative impact of noise and irrelevant content from other classes. We construct a memory bank to store the prototypes calculated in each FSL task and update them via a gated recurrent unit (GRU). Dynamically updating the prototypes can fuse the prototypes of multiple FSL tasks and enhance their representativeness in the feature space.

Finally, we input positive sample pairs into a multi-round reasoning module to learn semantic clues across dialogues, obtain reasoning features, feed them into a classifier for classification results and calculate multi-task loss. Experimental results show that the proposed method performs well on two public datasets and a private real-world dataset. The main contributions of this paper are as follows:

- A novel few-shot contrastive learning framework is designed for multi-round dialogue intent classification, including a pre-training stage and a contrastive meta-learning stage.
- A contrastive learning module is proposed with inference and context reasoning, which can alleviate the problem raised by word repetitions and unclear intentions within an ultra-long text content.
- A prototype update module (PUM) is introduced to rectify the biased prototypes by using GRU to update the prototypes stored in the memory bank and FSL task.

2 | Related Works

Few-shot text classification has garnered significant attention in recent years. Zhang et al. (2024) propose an LLM-augmented unsupervised contrastive learning framework for the few-shot text classification method, which introduces a cognition-enabled large language model (LLM) for efficient data augmentation and presents corresponding contrastive learning strategies. Liao et al. (2023) present a mask-guided BERT, a simple and modular framework to help BERT-based architectures tackle few-shot text classification. Loukas et al. (2023) propose using conversational GPT models for easy and quick few-shot text classification in the financial domain. Hou et al. (2021) present a FSL for multi-label intent classification, which introduces a metric learning-based method with anchored label representation. However, most of the research deals with IC problems generated in real customer service scenarios, which generate complex multi-round dialogues.

Contrastive learning has been extensively utilised in unsupervised representation learning by minimising the distance between positive pairs and maximising the distance between negative pairs. Many contrastive learning models incorporate components like detach, predictor, or intersection comparison (Cheng et al. 2023; Wu, Hovakimyan, and Hobbs 2023; Wang, Zhang, et al. 2023). He et al. (2023) propose an uncertainty estimation for few-shot text classification with contrastive learning, which can be trained with only one support sample for each class with the help of pseudo-uncertainty scores. Wang, Tan, et al. (2023) study the shortage issue of labels and create a novel contrastive meta-learning framework on graphs with a contrastive two-step optimisation and a similarity-sensitive mix-up strategy.

Intent Classification is essential in the natural language understanding module of question answering (QA). TextCNN uses convolution to capture n-grams in sentences (Wang et al. 2020) and captures essential features in the text through maximum pooling while solving the problem of inconsistent sentence length. Dialogue text classification technology has recently gained significant attention in academia and industry. The MFDG (Pang et al. 2022) model uses a heterogeneous graph-based encoder that achieves the whole dialogue by explicitly modelling multiple factors crucial for cross-lingual speaker-specific and contextual information extraction and can achieve a specific performance. Gong et al. (2023) propose a POS-aware adjacent relation attention network for question classification, which enhance context representations with POS information and neighbouring signals. The proposed adjacent relation attention mechanism can capture both the long-term dependency and local representation of semantic relations among words in different sentences. DialGNN (Yan et al. 2024) introduces a heterogeneous graph neural network framework tailored for the problem of dialogue classification, which exhibits a good performance in public China Mobile (CM) dataset with multi-round dialogue of formal language.

3 | The Proposed Method

The overview model architecture is shown in Figure 2, which mainly consists of a pre-training stage, a contrastive meta-training stage, a contrastive learning module, a multi-round scenario inference module and a loss function.

First, the input ASR text is divided into n rounds of dialogue based on the speakers in the conversation. To include connections between dialogues as much as possible in each round, each dialogue round includes multiple dialogues between [CLI, CSR, ...]. The total length of each dialogue round is L_{dia}, with each chunk size being Z, a hyperparameter. Therefore, the result of dividing a conversation text is n chunks, $U = [u_1, u_2, \ldots, u_n]$. Samples from China Mobile (CM) and China Telecom (CQTel.10000) are shown in Figure 5.

3.1 | Pre-Training and Prototype Update Module

As shown in Figure 2, a convolutional neural network classifier is constructed in the Pre-training stage and trained on the base dataset. Then, the last fully connection layer is removed to get the feature extractor f_{θ_f} with parameters θ_f. To train an efficient f_{θ_f} that offers good feature embeddings, the proposed method utilises the supervised contrastive loss to promote the classification performance and uses knowledge distillation to enable f_{θ_f} to effectively distinguish foreground objects and the background of the samples. Then, the feature extractor is transferred to subsequent stages. In the meta-training stage, we propose a Prototype Update Module (PUM) that constructs a memory bank to store the prototype of each class. The PUM uses the prototypes from each task to update the meta-knowledge of corresponding classes preserved in the memory bank. Meta-knowledge aids in learning new concepts from new ASR text samples. We

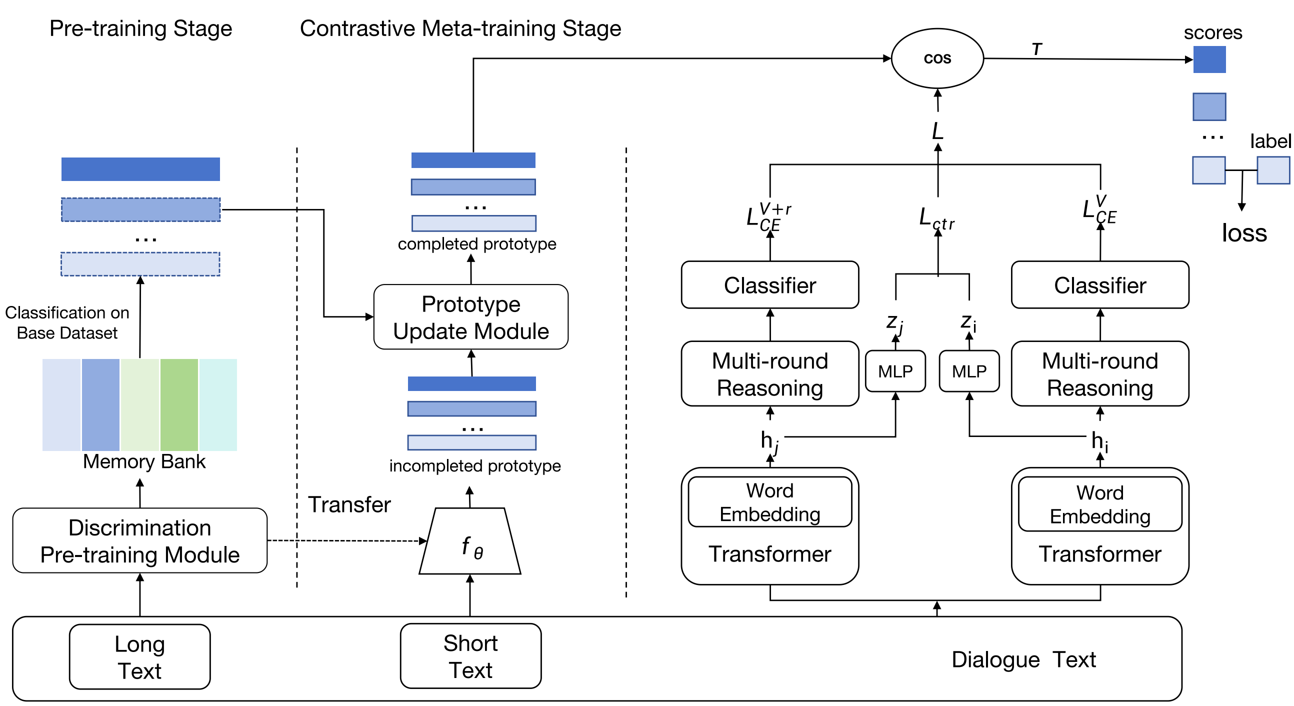

FIGURE 2 | The overview framework of the proposed method.

consider the prototype of each class in the base dataset C_{base} as meta-knowledge. Query samples are classified by calculating the Euclidean distance between their features and these prototypes. In the meta-test stage, query instance representations are compared with the prototypes in the memory bank to classify the query samples.

3.2 | Contrastive Learning Module

Adversarial samples are data that cause a model to misclassify after adding slight perturbations. Training on both refined and adversarial samples is an effective method to improve the robustness of the model. For a given loss function $\mathcal{L}\left(f_\theta\left(x_i+r\right), y_i\right)$, where f_θ is a parameterised neural network and x_i represents the input example, perturbations can be maximised by the loss function.

$$\max_r \ \mathcal{L}\left(f_\theta\left(x_i+r\right), y_i\right), \text{s. t. } |r|_\infty < \epsilon, \text{where } \epsilon > 0 \tag{1}$$

Using a first-order approximation, the loss function can be approximated as:

$$L\left(f_\theta\left(x_i+r\right), y_i\right) \approx L\left(f_\theta\left(x_i\right), y_i\right) + \nabla x_i L\left(f_\theta\left(x_i\right), y_i\right)^T r \tag{2}$$

By solving Equations (1 and 2), the perturbations can be derived in the following form:

$$r = -\epsilon \operatorname{sign}\left(\nabla_{x_i} \mathcal{L}\left(f_\theta\left(x_i\right), y_i\right)\right) \tag{3}$$

In this article, we generate adversarial samples by directly adding perturbations to the word embedding matrix in Transformer-based encoders. Specifically, after each forward pass on refined samples, we compute the gradient of the classifier loss to obtain the word embedding matrix V, which replaces the word embedding in Equation (3) to calculate perturbations. The loss function used by the classifier is the cross-entropy loss:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N} \sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \log\left(p\left(y_{i,c} \mid h^i_{[\text{CLS}]}\right)\right) \tag{4}$$

Based on the concept of contrastive learning, we treat the hidden vector $h^i_{[\text{CLS}]}$ from the last layer of the Transformer for a refined sample and its corresponding adversarial samples as a positive sample pair $h^j_{[\text{CLS}]}$. Then, we apply a non-linear projection layer:

$$z_i = W_2 \text{ReLU}\left(W_1 h^i_{[\text{CLS}]}\right) \tag{5}$$

$$z_j = W_2 \text{ReLU}\left(W_1 h^j_{[\text{CLS}]}\right) \tag{6}$$

Here, $W_1 \in R^{d_h \times d_h}$, $W_2 \in R^{d_k \times d_h}$, $d_k = 300$. For a batch containing N sample pairs, each positive sample pair has $2(N-1)$ negative sample pairs. The contrastive loss is defined as:

$$\mathcal{L}_{\text{ctr}} = -\log \frac{\exp\left(\text{sim}\left(z_i, z_j / \tau\right)\right)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp\left(\text{sim}\left(z_i, z_j / \tau\right)\right)} \tag{7}$$

where sim denotes the cosine similarity between two vectors, and τ is a hyperparameter.

3.3 | Multi-Round Inference Module

3.3.1 | Multi-Round Inference Module

As shown in Figure 3, the multi-round context inference module is composed of multi-round inference modules connected together. During the tth round of reasoning, we use an LSTM network to learn the intrinsic logical sequence and integrate contextual cues into the working memory, represented as:

$$\tilde{q}_i^{(t-1)}, h_i^{(t)} = \overrightarrow{\text{LSTM}}\left(q_i^{(t-1)}, h_i^{(t-1)}\right) \tag{8}$$

where $\tilde{q}_i^{(t-1)} \in R^{2d_u}$ is the output vector, and $q_i^{(t)} \in \mathbb{R}^{4d_u}$ is the context representation for the current round of dialogue. This context is initialised via c_i^q, for example, as $q_i^{(0)} = W_q c_i^q + b_q$, where $W_q \in \mathbb{R}^{4d_u \times 2d_u}$ and $b_q \in \mathbb{R}^{4d_v}$ are learnable parameters. $h_i^{(t)} \in \mathbb{R}^{2d_u}$ represents the working memory, which can store and update previous memories $h_i^{(t-1)}$. Through iterative updates of the working memory, the implicit logical sequence among cues can be learned. The initial state $h_i^{(t)}$ is set to zero, and t represents the index for computing the final state.

3.3.2 | Context Reasoning Module

For the retrieval process, we use an attention mechanism to match relevant contextual cues from global memory in Figure 4. The detailed computation is as follows:

$$e_{ij}^{(t-1)} = f\left(g_j, \widetilde{q}_i^{(t-1)}\right) \tag{9}$$

$$\alpha_{ij}^{(t-1)} = \frac{\exp\left(e_{ij}^{(t-1)}\right)}{\sum_{j=1}^{N} \exp\left(e_{ij}^{(t-1)}\right)} \tag{10}$$

$$r_i^{(t-1)} = \sum_{j=1}^{N} \alpha_{ij}^{(t-1)} g_j \tag{11}$$

where f is a function (e.g., dot product) calculating a scalar from g_j and $\widetilde{q}_i^{(t-1)}$. Here, g_j is obtained by concatenating the first j slices processed through a Transformer.

Afterward, $\widetilde{q}_i^{(t-1)}$ and $r_i^{(t-1)}$ are concatenated to form the next input $q_i^{(t)}$:

$$q_i^{(t)} = \left[\widetilde{q}_i^{(t-1)}; r_i^{(t-1)}\right] \tag{12}$$

The query $q_i^{(t)}$ is updated through the working memory $h_i^{(t)}$, allowing it to retrieve more contextual cues.

For a given dialogue segment context representation u_i, global memory G^q, and number of iterations T^q, the entire retrieval phase can be represented as $q_i{}^q = \text{Cognition}\left(c_i{}^q, G^q; T^q\right)$.

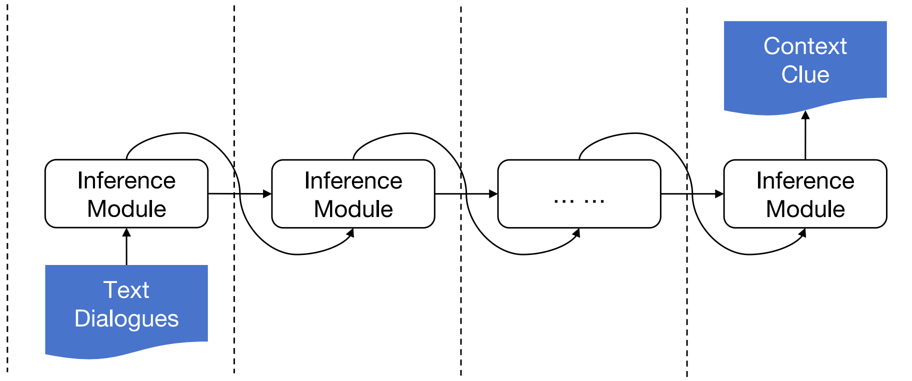

FIGURE 3 | The multi-round inference module.

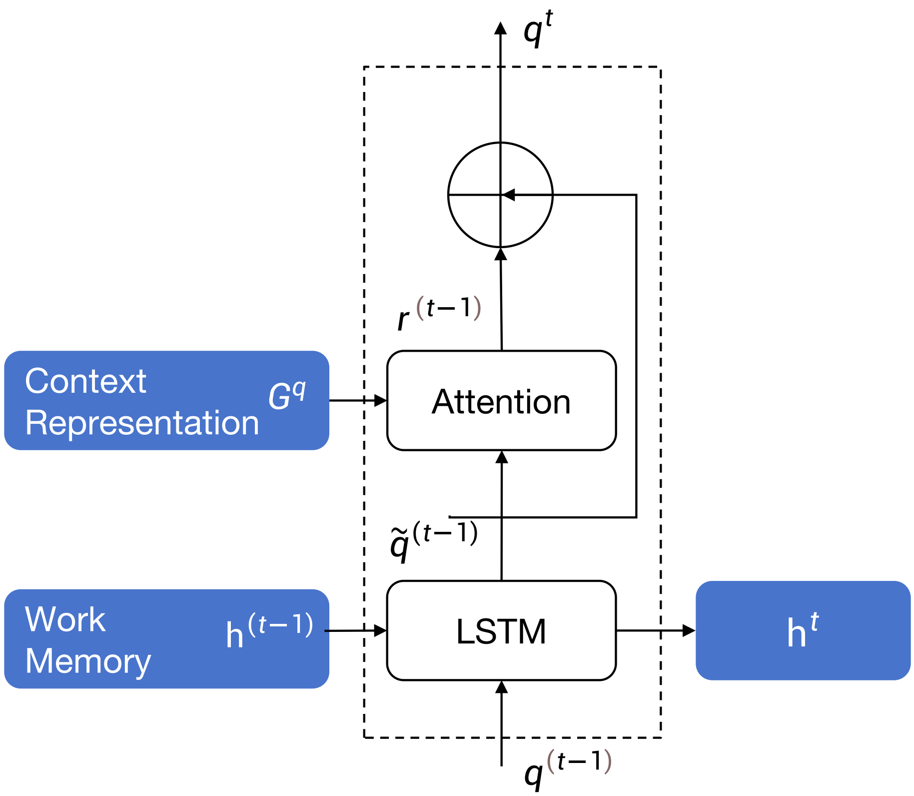

FIGURE 4 | The context reasoning module.

To ensure that the features $Q_j{}^q = \left[q_1^q, q_2^q, \ \ldots \ , q_n^q\right]$ extracted from the jth dialogue text represent all semantic information, we choose the maximum number of dialogue rounds n as the length of feature Q. For texts with fewer than n rounds of dialogue, we pad with zeros.

3.3.3 | Loss Function

The total loss function is as follows, using a multi-task approach that combines classification loss and contrastive loss with weighted operations:

$$\mathcal{L} = \frac{(1-\lambda)}{2}\left(\mathcal{L}_{\mathrm{CE}}^{V} + \mathcal{L}_{\mathrm{CE}}^{V+r}\right) + \lambda \mathcal{L}_{\mathrm{ctr}} \tag{13}$$

where $\mathcal{L}_{\mathrm{CE}}^{V}$ is the classification loss from refined samples, $\mathcal{L}_{\mathrm{CE}}^{V+r}$ is the classification loss from adversarial samples, and $\mathcal{L}_{\mathrm{ctr}}$ is the contrastive loss.

4 | Experimental Evaluations

4.1 | Datasets and Implementations

We evaluate the proposed method on three datasets: the public available multi-round dialogue dataset DailyDialog (Li et al. 2017) and phone call dialogues from China Mobile(CM) (Yan et al. 2024), and a private ASR text dataset from China Telecom Chongqing Branch(CQTel.10000). The CQTel.10000 is similar with CM, which comprises 80,438 entries, categorised into 5 major types:

Example from China Telecom（CQTel.10000）

... ...

CLI: Hi, I want to ask why we can't use this internet broadband?(你好我想问一下为什么我们这个宽带用不了呀？)

CSR: Internet for this number?(本机号码的宽带吗？)

CSR: Okay, just a minute. I'll check.（好的稍等一下我帮您查询，）

CSR: Are there any flashing red lights on the modem and router.(光猫和路由器有没有闪红灯呢。)

CLI: Modem is red.(光， 毛鸿，)

... ...

CLI: Okay, and then there's another question I'd like to find out how much I'm spending per month.(好的然后还有一个问题我想查一下我这个每个月消费是多少啊。)

CSR: Wait a moment.(稍等一下？)

CSR: It's a fixed package, 1199 RMB Business Enterprise Pass package.(它这个固定的资费的话都比较高是一个幺幺九九的一个商企通套餐。)

CLI: I know ah but then just I called in he said I consumed 217 RMB, then how do I consume 217RMB considering the 199 package. I never make call over limitation and do not use internet.(我知道啊可是当时刚刚我打电话进来他说我消费了二百一十七，那么我怎么消费二百一十七那我这个就是幺九九包含的一个套餐呢，我话费没有打超过我流量也没用，如果多然后为什么买来的二百一期呢？)

Complex, Informal Language, Multi-intent, Repetition, etc.

Example from China Mobile（CM, DialGNN）

CLI: I'm asking if you could send the electronic invoice to my email. I haven't received the invoices for June and July. (我想请您把电子发票发到我的邮箱。我还没有收到6月和7月的发票)

CSR: If you're referring to Joly and you have a postpaid number, you need to settle your bill first. After that, it takes about three days fro you to be able to receive our invoice. However, your July bill hasn't been settled.(如果您指的是7月份，并且您有一个后付费号码，您需要先结清账单。之后，您需要三天左右的时间才能收到我们的发票。但是，您的七月账单尚未结清)

CSR: I checked the status of your July bill, and it hasn't been settled, ma'am.(我查了你七月份的账单，还没有结清，女士)

CLI: What is the specific amount?(具体的金额是？)

CSR: Please wait a moment while I check. Ah, for June, the total is 124.31 RMB.(请稍等片刻，我正在核对。6月份的总额是124.31元人民币)

Simple, Formal Language, Single-intent

FIGURE 5 | The examples of case study on CQTel.10000 and CM datasets.

TABLE 1 | Experiments on DailyDialog, CM and CQTel.10000.

Datasets	Model	Acc	Top-3	Top-5	*F1*
DailyDialog	TextRNN	49.10	74.90	74.00	38.62
	Bert-base (Wang et al. 2020)	62.90	83.40	88.60	57.48
	DialogueGCN (Ghosal et al. 2019)	61.30	83.90	90.80	52.48
	RGAT (Schlichtkrull et al. 2018)	63.50	89.40	93.90	59.02
	DAG (Yu et al. 2021)	63.20	86.40	93.20	58.52
	MFDG (Pang et al. 2022)	**66.50**	89.30	94.40	60.64
	Ours	65.98	**89.55**	**95.10**	**61.23**
CM	DialGNN(CNN_LSTM) (Yan et al. 2024)	54.10	87.23	89.95	41.80
	DialGNN(Han) (Yan et al. 2024)	63.10	90.07	91.28	48.10
	DialGNN(BERT) (Yan et al. 2024)	70.20	92.26	96.47	59.30
	Ours	**71.31**	**94.74**	**96.47**	**61.17**
CQTel.10000	DialGNN(BERT) (Yan et al. 2024)	71.96	80.27	83.11	58.31
	Ours	**76.22**	**83.36**	**85.48**	**68.72**

Note: Bold indicates best Results.

enquiry (查询), *processing* (办理), *consultation* (咨询), *complaint* (投诉) and *fault report* (报修), with the top 20 subcategories selected under each major category. As shown in Figure 5, the ASR text in CQTel.10000 is more complex compared with CM dataset. For example, the text is not formal language, and many mistakes contains (e.g., 1199 RMB should be 199RMB). The critical issue is that there may exists many intents(blue for *Fault Report*, yellow for *Enquiry*) in the dialogues, which is normal in the real environment, however, this is few in CM.

The experiments were conducted using a V100 GPU (32 GB), implemented based on the Pytorch. The initial learning rate of the model is set to 1e-4, and the Adam optimiser is used during the training phase. The batch size is set to 32 and the results are averaged over five experiments, and each training session consisting of 60 iterations. We use ResNet-12 as a feature extractor and a list of 64×128 dimensions is used as the memory bank. Moreover, a two-layer GRU with a 512-dimensional hidden layer contributes to the prototype update. Both datasets are split into training and testing sets in an 8:2 ratio. We follow the evaluation metrics in (Yan et al. 2024): accuracy (Acc), precision (P), recall (R) and $F1$-score.

4.2 | Comparison With State-Of-The-Art Methods

As shown in Table 1, it is obvious that the proposed method outperforms most of the methods and achieves similar results as MFDG on DailyDialog. It performs better in ASR text classification in telecommunication scenarios. The proposed method exhibits significant improvement over DialGNN on CM and CQTel.10000 datasets. The proposed method shows better performance in the multi-round intent classification considering informal language (e.g., repetition, inconsistency, multi-intents).

Then, we conducts experiments and comparative analysis on CQTel.10000 dataset with all 5 business types, introduces Top k accuracy assessment method and three indicators, Top 2, Top 3 and Top 5, and the experimental results are shown in Table 2. The proposed method exhibits better Acc in *processing* (办理), which indicates that user prefer to express their intent directly. However, when client enquiry some information, they may transfer to other intents (complex intents in one dialogue, examples in Figure 5) which leads to a decrease in Acc.

We evaluate the proposed method with other FSL or few-shot contrastive learning methods for intent classification on three datasets. As shown in Table 3, the proposed method outperforms other method in all 5-way 1-shot and 5-way 5-shot.

4.3 | Ablation Studies

In this section, we conducted extensive ablation experiments on the proposed modules to validate their effectiveness. The results are presented in Table 4, which demonstrate that the complete model achieved optimal performance. Furthermore, all modules contributed to the performance improvement to some extent. Experimental results demonstrate that the prototype update module (PUM), which constructs a memory bank and imports a gated recurrent unit, plays a more significant role than other components in existing modules.

TABLE 2 | Experimental evaluation on CQTel.10000 by bussiness types.

Types	P(Top1)	R	$F1$	Top2	Top3	Top5	Acc	Support
Processing (办理)	0.820	0.811	0.813	0.932	0.968	0.985	0.890	1132
Enquiry (查询)	0.653	0.649	0.649	0.749	0.788	0.819	0.680	2029
Fault Report (报修)	0.689	0.666	0.672	0.795	0.841	0.855	0.756	1691
Complaint (投诉)	0.620	0.608	0.610	0.688	0.735	0.748	0.750	3320
Consultation (咨询)	0.708	0.688	0.692	0.803	0.836	0.867	0.735	1536
Total	0.698	0.684	0.687	0.793	0.834	0.855	0.762	9708

TABLE 3 | The few-shot IC results (5-way).

	DailyDialog		CQTel.10000		CM	
Methods	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
MAML (Finn, Abbeel, and Levine 2017)	30.27	34.63	32.19	36.43	39.13	41.27
DHLNet (Zhang et al. 2022)	54.23	55.16	60.99	64.17	62.22	66.05
LA-UCL (Zhang et al. 2024)	59.06	61.05	65.48	71.99	69.74	73.32
Ours	64.26	66.85	66.04	72.47	73.18	78.29

TABLE 4 | Ablation studies (5-way).

	DailyDialog		CQTel.10000		CM	
Module	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Full	64.26	66.85	66.04	72.47	73.18	78.29
SupLoss	63.87	66.02	65.84	71.88	71.66	77.01
DPM	62.99	64.28	65.01	70.65	71.13	75.87
PUM	61.34	63.53	64.27	69.28	68.34	71.26

5 | Conlusion

In this article, we effectively solve the problem of intent classification in multi-round dialogues and improve the robustness of the model based on a few-shot contrastive learning method. The experimental evaluation demonstrates that the proposed method outperform stat-of-the-art methods in both public datasets and private dataset. The proposed method has been implemented in China Telecom Chongqing for 10,000 service centre to quickly classify the user's intention to understand customers.

Acknowledgements

This research was supported in part by the Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant No. KJZD-M202400603), in part by the Key Cooperation Project of Chongqing Municipal Education Commission (No. HZ2021008) and in part by the Project of Key Laboratory of Tourism Multisource Data Perception and Decision, Ministry of Culture and Tourism, China (No. H2023009).

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Cheng, T., H. Cheng, Y. Fang, Y. Liu, and C. Gao. 2023. "CPCL: Conceptual Prototypical Contrastive Learning for Few-Shot Text Classification." *Journal of Intelligent & Fuzzy Systems* 45, no. 6: 11963–11975.

Duran, N., S. Battle, and J. Smith. 2023. "Sentence Encoding for Dialogue Act Classification." *Natural Language Engineering* 29, no. 3: 794–823.

Finn, C., P. Abbeel, and S. Levine. 2017. "Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks." In *International Conference on Machine Learning*, 1126–1135. PMLR.

Firdaus, M., H. Golchha, A. Ekbal, and P. Bhattacharyya. 2021. "A Deep Multi-Task Model for Dialogue Act Classification, Intent Detection and Slot Filling." *Cognitive Computation* 13: 626–645.

Ghosal, D., N. Majumder, S. Poria, N. Chhaya, and A. Gelbukh. 2019. "DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation." In *Proceedings of EMNLP-IJCNLP*, 154–164.

Gong, P., J. Liu, Y. Xie, M. Liu, and X. Zhang. 2023. "Enhancing Context Representations With Part-Of-Speech Information and Neighboring Signals for Question Classification." *Complex & Intelligent Systems* 9, no. 6: 6191–6209.

Hayat, S., A. Rextin, A. Idris, and M. Nasim. 2020. "Text and Phone Calls: User Behaviour and Dual-Channel Communication Prediction." *Human-centric Computing and Information Sciences* 10, no. 1: 11.

He, J., X. Zhang, S. Lei, et al. 2023. "CLUR: Uncertainty Estimation for Few-Shot Text Classification With Contrastive Learning." In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 698–710.

He, Y., and S. Young. 2003. "A Data-Driven Spoken Language Understanding System." In *IEEE Workshop on Automatic Speech Recognition and Understanding*, 583–588.

Hou, Y., Y. Lai, Y. Wu, W. Che, and T. Liu. 2021. "Few-Shot Learning for Multi-Label Intent Detection." *Proceedings of the AAAI Conference on Artificial Intelligence* 35, no. 14: 13036–13044.

Li, J. 2022. "Multi-Round Dialogue Intention Recognition Method for a Chatbot Baed on Deep Learning." In *International Conference on Multimedia Technology and Enhanced Learning*, 561–572. Cham, Switzerland: Springer Nature Switzerland.

Li, Y., H. Su, X. Shen, W. Li, Z. Cao, and S. Niu. 2017. "DailyDialog: A Manually Labelled Multi-Turn Dialogue Dataset." In *Proceedings of the International Joint Conference on Natural Language Processing*, 986–995.

Liao, W., Z. Liu, H. Dai, et al. 2023. "Mask-Guided Bert for Few Shot Text Classification." arXiv preprint arXiv:2302.10447.

Liu, H., F. Zhang, X. Zhang, S. Zhao, and X. Zhang. 2021. "An Explicit-Joint and Supervised-Contrastive Learning Framework for Few-Shot Intent Classification and Slot Filling." arXiv preprint arXiv:2110.13691.

Liu, X., A. Eshghi, P. Swietojanski, and V. Rieser. 2021. "Benchmarking Natural Language Understanding Services for Building Conversational Agents." In *Increasing Naturalness and Flexibility in Spoken Dialogue Interaction: 10th International Workshop on Spoken Dialogue Systems*, 165–183. Springer Singapore.

Loukas, L., I. Stogiannidis, P. Malakasiotis, and S. Vassos. 2023. "Breaking the Bank With Chatgpt: Few-Shot Text Classification for Finance." arXiv preprint arXiv:2308.14634.

Louvan, S., and B. Magnini. 2020. "Recent Neural Methods on Slot Filling and Intent Classification for Task-Oriented Dialogue Systems: A Survey." arXiv Preprint arXiv:2011.00564.

Madotto, A., Z. Liu, Z. Lin, and P. Fung. 2020. "Language Models as Few-Shot Learner for Task-Oriented Dialogue Systems." arXiv Preprint arXiv:2008.06239.

Pang, J., H. Xu, S. Song, B. Zou, and X. He. 2022. "Mfdg: A Multi-Factor Dialogue Graph Model for Dialogue Intent Classification." In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases,* 691–706. Cham, Switzerland: Springer International Publishing.

Sauer, A., S. Asaadi, and F. Küch. 2022, May. "Knowledge Distillation Meets Few-Shot Learning: An Approach for Few-Shot Intent Classification Within and Across Domains." In *Proceedings of the 4th Workshop on NLP for Conversational AI*, 108–119.

Schlichtkrull, M., T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling. 2018. "Modeling Relational Data With Graph Convolutional Networks." In *The Semantic Web: 15th International Conference*, 593–607. Heraklion, Greece: ESWC.

Tian, J., Z. Tu, N. Li, T. Su, X. Xu, and Z. Wang. 2022. "Intention Model Based Multi-Round Dialogue Strategies for Conversational AI Bots." *Applied Intelligence* 52, no. 12: 13916–13940.

Wang, H., J. He, X. Zhang, and S. Liu. 2020. "A Short Text Classification Method Based on N-Gram and CNN." *Chinese Journal of Electronics* 29, no. 2: 248–254.

Wang, X., S. Zhang, Z. Qing, et al. 2023. "Molo: Motion-Augmented Long-Short Contrastive Learning for Few-Shot Action Recognition." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18011–18021.

Wang, S., Z. Tan, H. Liu, and J. Li. 2023. "Contrastive Meta-Learning for Few-shot Node Classification." In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2386–2397.

Wu, J., N. Hovakimyan, and J. Hobbs. 2023. "Genco: An Auxiliary Generator From Contrastive Learning for Enhanced Few-Shot Learning in Remote Sensing." arXiv preprint arXiv:2307.14612.

Xiang, L., J. Qin, X. Xiang, and Y. Tan. 2024. "Text Coverless Information Hiding Based on BERT Entity Recognition." *Human-centric Computing and Information Sciences* 14, no. 31: 1–16.

Xie, Q., T. Zhu, J. Wang, L. L. Chen, C. Liu, and Y. Wan. 2024. "A Two-Step Channel Mask Method for Data Augmentation in Contrastive Learning-Based Human Activity Recognition." *Human-centric Computing and Information Sciences* 14, no. 26: 1–16.

Xu, H., J. Pang, S. Song, and B. Zou. 2023. "Improving Dialogue Intent Classification With a Knowledge-Enhanced Multifactor Graph Model (Student Abstract)." *Proceedings of the AAAI Conference on Artificial Intelligence* 37, no. 13: 16366–16367.

Yan, Y., B. W. Zhang, P. H. Min, G. W. Ding, and J. Y. Liu. 2024. "DialGNN: Heterogeneous Graph Neural Networks for Dialogue Classification." *Neural Processing Letters* 56, no. 2: 143.

Yu, D., Y. Yang, R. Zhang, and Y. Wu. 2021. "Knowledge Embedding Based Graph Convolutional Network." In *Proceedings of the Web Conference*, 1619–1628.

Yuan, W. 2023. "Few-Shot User Intent Detection and Response Selection for Conversational Dialogue System Using Deep Learning." https://hdl.handle.net/10315/41378.

Zhang, J., H. Gao, P. Zhang, B. Feng, W. Deng, and Y. Hou. 2024. "LA-UCL: LLM-Augmented Unsupervised Contrastive Learning Framework for Few-Shot Text Classification." In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation*, 10198–10207.

Zhang, X., Y. Zhang, Z. Zhang, and J. Liu. 2022. "Discriminative Learning of Imaginary Data for Few-Shot Classification." *Neurocomputing* 467: 406–417.

Zhao, D., D. Han, Y. Yuan, C. Wang, and S. Song. 2023. "MuSE: A Multi-Scale Emotional Flow Graph Model for Empathetic Dialogue Generation." In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 491–507. Cham, Switzerland: Springer Nature Switzerland.