Pseudo-Label Enhanced Prototypical Contrastive Learning for Uniformed Intent Discovery

Anonymous ACL submission

Abstract

New intent discovery is a crucial capability for task-oriented dialogue systems. Existing methods focus on transferring in-domain (IND) prior knowledge to out-of-domain (OOD) data through pre-training and clustering stages. They either handle the two processes in a pipeline manner, which exhibits a gap between intent representation and clustering process or use typical contrastive clustering that overlooks the potential supervised signals from the whole data. Besides, they often deal with either open intent discovery or OOD settings individually. To this end, we propose a Pseudo-Label enhanced Prototypical Contrastive Learning (PLPCL) model for uniformed intent discovery. We iteratively utilize pseudo-labels to explore potential positive/negative samples for contrastive learning and bridge the gap between representation and clustering. To enable better knowledge transfer, we design a prototype learning method integrating the supervised and pseudo signals from IND and OOD samples. In addition, our method has been proven effective in two different settings of discovering new intents. Experiments on two benchmark datasets and two task settings demonstrate the effectiveness of our approach.[1]

1 Introduction

New intent discovery, aiming to uncover and categorize out-of-domain intents absent in training data has received increasing attention due to its crucial role in dialogue systems (Min et al., 2020; Vedula et al., 2019). Initially, researchers focused on exploring unsupervised clustering methods (Hakkani-Tür et al., 2015; Liu et al., 2021; Shi et al., 2018). However, real-world scenarios typically involve limited labeled data, prompting a shift to semi-supervised approaches, notably OOD and open intent discovery (Lin et al., 2020; Mou et al., 2022b,c; Shen et al., 2021; Zhang et al., 2021b, 2022). OOD intent discovery clusters solely unlabeled OOD intents using labeled IND data, while open intent discovery aims to recognize both known and new categories from unlabeled data (see Figure 1).

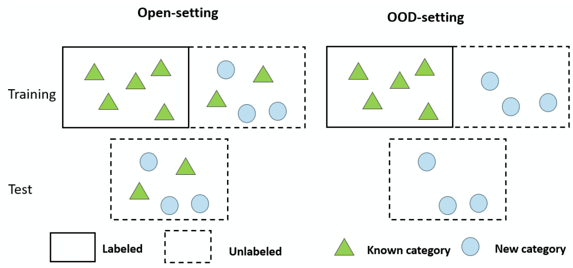


Figure 1: Two basic task setting for intent discovery. In open-setting: Partially labeled IND data for training, all categories in test (IND and OOD). In OOD setting: Fully labeled IND data for training, only OOD categories in test.

The existing approaches for new intent discovery typically use a two-stage contrastive clustering approach, involving IND pretraining and OOD clustering. For open intent discovery, researchers focus on effectively utilizing the small amount of labeled data for weakly supervised (Lin et al., 2020) or semi-supervised clustering methods (Shen et al., 2021; Zhang et al., 2021b). For OOD intent discovery, previous works commonly employ a contrastive clustering framework (Li et al., 2021) with approaches such as multi-head contrast learning (Mou et al., 2022b) or neighbor-enhanced contrastive strategies (Mou et al., 2022c).

Although previous methods in intent discovery have achieved notable success, several challenges in the field remain unexplored. (1) One key challenge is efficiently integrating labeled and unlabeled data for joint representation learning and clustering. While some methods focus on contrastive clustering for joint learning (Mou et al., 2022b), they often overlook critical supervised signals from IND samples during the clustering stage.

[1]The codes and datasets will be publicly accessible upon acceptance

Leveraging labeled information for this process remains underexplored, particularly given the limited availability of labeled data and the risk of overfitting. (2) Another critical issue is devising effective transfer learning mechanisms between IND and OOD data while preventing catastrophic forgetting. Existing methods often discard classifiers trained on prior knowledge, retaining only feature extraction during clustering with OOD samples (Zhang et al., 2021b). This requires additional alignment strategies, potentially introducing noise if suboptimal. There's an urgent need for methods preserving prior knowledge while adapting better to new intent data, ensuring seamless transfer learning. (3) Moreover, the predominant focus of prior research has been either on open intent discovery or OOD intent discovery individually, disregarding the practical need for a unified method capable of handling both scenarios (Mou et al., 2022b; Zhang et al., 2021b). Real-world dialogue systems frequently encounter situations necessitating updates or migration, highlighting the imperative for a uniform intent discovery approach adaptable to varying system changes.

To address these limitations, we propose a **P**seudo-**L**abel enhanced **P**rotypical **C**ontrastive **L**earning (PLPCL) model which is built upon contrastive clustering for joint representation learning and clustering. Our approach begins by pretraining the contrastive clustering model using labeled IND data. To effectively harness labeled information, we integrate IND samples with unlabeled data using a semi-supervised clustering strategy, employing distinct contrastive learning strategies for labeled and unlabeled data. To prevent overfitting and maximize the utilization of unlabeled data, we iteratively select reliable unlabeled samples with confident pseudo-labels. These reliable samples serve as potential positive/negative samples during contrastive learning, enhancing the overall contrastive clustering process.

To bridge the gap between IND and OOD data, we introduce a prototype learning strategy. It keeps the prototype matrix by integrating instance and cluster features from both the IND and reliable unlabeled samples. Our method seamlessly integrates contrastive clustering and prototypical learning, eliminating the need for an extra aligning module. This integration facilitates improved knowledge transfer from IND to OOD without discarding label information from IND samples. Furthermore, our framework is purposefully designed for uniform intent discovery, demonstrating effectiveness in both open-setting and OOD-setting scenarios.

The contribution of our work is threefold:

- We introduce a novel method that leverages both labeled and unlabeled data through pseudo-label enhanced semi-supervised contrastive learning. This approach facilitates joint representation learning and clustering by effectively leveraging the whole data.
- We propose a prototypical contrastive learning framework for uniformed intent discovery integrating prototypical learning and contrastive clustering utilizing labeled and unlabeled samples to bridge the gap between IND prior knowledge and OOD categories.
- We conduct extensive experiments on both OOD intent discovery and open intent discovery scenarios and the results demonstrate the effectiveness of our proposed method.

2 Related Work

2.1 Intent Discovery.

Recent research for intent discovery can be broadly categorized into OOD-setting and open-setting. As shown in Figure 1, open intent discovery involves clustering both IND and OOD intents with IND priori knowledge. Samples with IND intents are not all labeled. Whereas OOD intent discovery focuses on accurately handling OOD intents and assumes that the intents of labeled and unlabeled data do not overlap, which means all IND samples are labeled. (Lin et al., 2020) proposed a self-supervised clustering method that utilizes limited labeled data. (Zhang et al., 2021b) proposed a k-means-based semi-supervised clustering method that can effectively use prior knowledge in intent discovery. (Mou et al., 2022b) proposed a disentangled contrastive learning framework that mainly focuses on OOD intent clustering and decouples instance and cluster-level features to disentangle the knowledge of IND and OOD samples. (Han et al., 2019) extended deep embedded clustering to transfer learning setting, incorporating prior knowledge for OOD clustering.

2.2 Contrastive Clustering

Contrastive clustering has been widely used in various clustering scenarios, such as unsupervised semantic segmentation (Hamilton et al., 2022) and generalized self-supervised contrastive learning (Hu et al., 2022). It has also been applied in OOD intent detection and discovery tasks (Kumar et al., 2022; Mou et al., 2022b,c). (Li et al., 2021) proposed a contrastive clustering framework with two contrastive learning heads. It provided objective guidance for clustering, avoiding interference from prior knowledge. (Mou et al., 2022b) extended contrast clustering to the semi-supervised scenario and designed a two-stage contrastive learning process that includes both supervised pre-training and unsupervised clustering. It achieved state-of-the-art results for OOD intent discovery.

2.3 Prototype Learning.

The prototype learning method is widely used in clustering analysis and classification problems. In semi-supervised clustering scenarios, coarsely assigned pseudo-labels may result in mismatches between instances and prototypes, introducing noise that significantly affects the clustering performance. (An et al., 2023) used weighted pseudo-labels to reduce the effect of mismatched prototypes. (Huang et al., 2022) proposed the approach of prototype scattering, which enhances the variance between the clusters by maximizing the distances between prototype features, to obtain well-separated clusters. The prototype learning method is robust to noise and outliers. Compared to other clustering methods, it is an intuitive and interpretable approach that can provide references for the entire cluster based on representative examples.

3 Preliminaries

First, we define the problem of uniformed intent discovery. Then we briefly introduce the contrastive clustering framework.

3.1 Problem Statement

OOD intent discovery suppose we have a set of labeled IND data D_{IND} and unlabeled OOD data D_{OOD}, and aim at clustering OOD intents. Note that there is no overlap between the IND data and the OOD data. Indeed, extending the OOD classifier to an all-category classifier is the extreme case of open intent discovery. Open intent discovery assumes that we have an intent analysis dataset D_l, D_u, where$D_l = \{(x_l, y_l)|y_l \in \mathcal{Y}_k\}$,$D_u = \{(x_u, y_u)|y_u \in \mathcal{Y}_k, \mathcal{Y}_{uk}\}$. $\mathcal{Y}_k$ are known intents and $\mathcal{Y}_{uk}$ are unknown intents. In the extreme case, all the samples of the known categories in the training set are labeled, and only data with unknown intents are contained in D_u. Then the prior labeling information is the same as the OOD classification. Since we do not have a priori assumptions that $y_l \cap y_u = \emptyset$, the setting is different from the generalized intent discovery (Mou et al., 2022a). In fact, we also compared the results under different labeled ratios with the same IND intent division.

Uniformed intent discovery includes both OOD intent discovery and open intent discovery. On the one hand, we consider the open setting when IND and OOD intents are not separated, i.e., when new intent detection needs to be introduced. On the other hand, we consider merely OOD intent clustering. This helps the system to adapt or transform to different scales of change.

3.2 Contrastive Clustering

Our model is based on a contrastive clustering framework. It performs instance-level and cluster-level contrastive learning. Specifically, for a given dataset, positive and negative instance pairs are constructed through data augmentation and then projected into a feature space. Instance-level and cluster-level contrastive learning are performed in the row space and column space, respectively.

Unsupervised instance-level contrastive learning (ILCL) is performed on unlabeled data, where the augmented sample of each sample is considered as a positive sample and other samples are considered as negative samples. f_i,f_j refer to the augmented samples that are generated from the same samples after passing through the dropout layer.

$$\ell_{i,j}^{ins} = -log\frac{\exp\left(sim(f_i, f_j)/\tau\right)}{\sum_{k=1}^{2N} 1_{i\neq k} \exp\left(sim(f_i, f_k)/\tau\right)} \tag{1}$$

On the cluster-level contrastive learning head g, it performs cluster-level contrastive learning (CLCL). The cluster representation of the augment sample is considered as a positive sample, and the other cluster representations are considered as negative samples. y_i refers to the representations of the clusters, which are columns in the cluster-level feature matrix. y_j are the dropout-augment representations for the cluster level.

$$\ell_{i,j}^{clu} = -log\frac{\exp\left(sim(y_i, y_j)/\tau\right)}{\sum_{k=1}^{2N} 1_{i\neq k} \exp\left(sim(y_i, y_k)/\tau\right)} \tag{2}$$

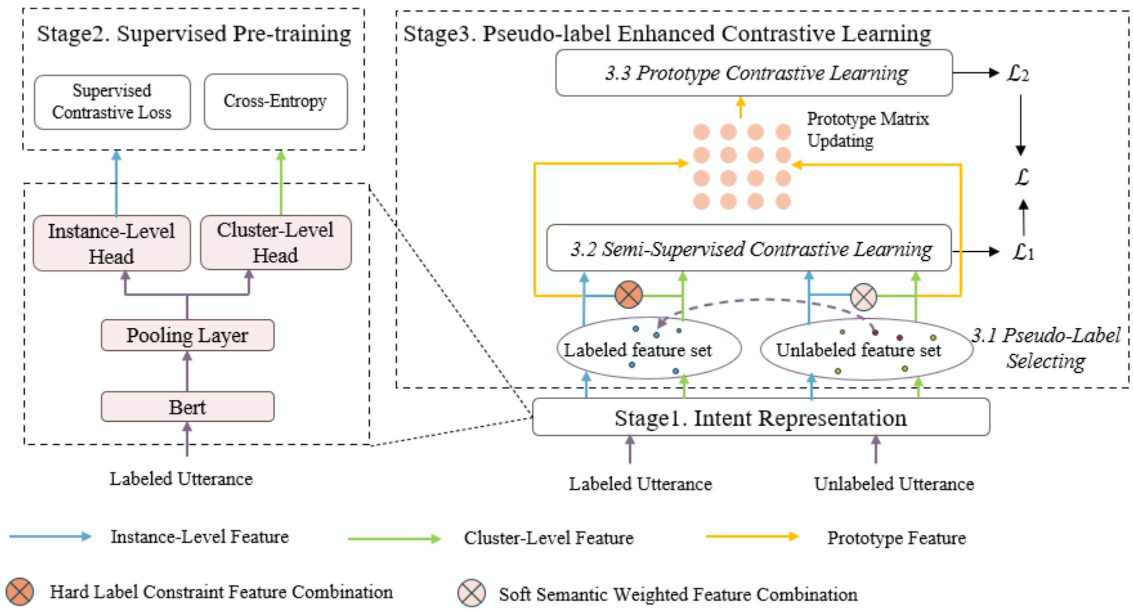


Figure 2: The overall architecture of the proposed PLPCL. Instance-Level Head and Cluster-Level Head are two separate MLPs. The pre-train stage involves labeled IND data. Through confident pseudo-label filtering, the whole training dataset is divided into supervised and unsupervised parts, which are then subjected to different contrastive learning strategies. Prototype features are computed based on instance-level features and cluster-level features.

4 Methods

Overall Architecture: The overall framework of PLPCL is illustrated in Figure 2. Our framework follows a two-stage pipeline. In the pre-train stage, IND-labeled samples are utilized for supervised multi-head contrastive learning to acquire prior knowledge. This model is adapted from contrastive clustering and includes intent representation alongside two independent heads. These heads are instrumental in decoupling the representation into instance-level and cluster-level spaces, facilitating joint representation learning and clustering.

After pretraining, the prototypes of known categories are obtained, serving as a foundation for efficient knowledge transfer across IND and OOD data. In the second stage, the multi-head contrastive model is further trained on the entire dataset including IND and OOD samples. Specifically, this stage comprises three iterative steps: pseudo-label selecting, semi-supervised contrastive learning, and prototype contrastive learning. These steps collectively aim to transfer prior knowledge to new categories and enhance the model's adaptability.

4.1 Intent Representation

To facilitate effective knowledge transfer between IND and OOD samples, we aim to achieve joint intent representation and clustering by learning instance-level and cluster-level features for each sample.

Drawing inspiration from (Mou et al., 2022b), we first extract the intent representation using a pre-trained BERT model and a pooling layer to extract text representation. Then we utilize two independent MLPs to map the intent representation z_i into two disentangled latent vectors: $f_i = f(z_i)$ and $g_i = g(z_i)$.

4.2 Supervised Pre-training

To familiarize the model with the prior knowledge obtained from labeled IND samples and to establish initial cluster prototypes, we conduct pre-training on IND samples. Based on the multi-level intent representation, we conduct two-level pertaining.

For instance-level representation, we adopt supervised contrastive learning (SCL) to maximize inter-class variance and minimize intra-class variance within the IND samples.

Formally, for a sample x_i in a mini-batch of size N, the samples within N sharing the same label are considered as positive samples, while the remainder is treated as negative samples. The SCL loss is computed as follows:

$$\mathcal{L}_{SCL} = \sum_{i=1}^{N} -\frac{1}{|N_{y_i} - 1|} \sum_{j=1}^{N} \mathbb{1}_{i \neq j} \mathbb{1}_{y_i = y_j} \log \frac{\exp\left(f_i \cdot f_j / \tau\right)}{\sum_{k=1}^{N} \mathbb{1}_{i \neq k} \exp\left(f_i \cdot f_k / \tau\right)} \tag{3}$$

where y_i, y_j are the labels of samples x_i, x_j and $\mathbb{1}$ is an indicator function. N_{y_i} denotes the number of samples in N with the label y_i. f_i, f_j indicate the instance-level representation of x_i, x_j. t is the temperature parameter for contrastive learning.

For the cluster-level representation, we apply cross-entropy loss (CE) to learn cluster-friendly features. Note that we use a classifier with both IND and OOD classes for open-setting to better reserve the priori knowledge extracted at this stage, allowing our model to better retain the prior knowledge acquired during the pre-training stage compared to previous works.

4.3 Semi-supervised Training

After pretraining with labeled data, we achieve a good initialization of representation learning and clustering. The challenge now lies in transferring the prior knowledge to new intents. There are two critical problems to be addressed: (1) Effectively utilizing both labeled and unlabeled data to enhance the joint representation and clustering process; (2) Transferring learned knowledge from IND to OOD data while continually refining representation learning to enhance cluster-friendly features without encountering catastrophic forgetting. To tackle these challenges, we introduce a pseudo-label enhanced contrastive learning scheme tailored for iterative clustering and updating. This scheme starts from reliable pseudo-label filtering for unlabeled samples, followed by semi-supervised contrastive learning and prototypical contrastive learning.

4.3.1 Pseudo-label Selecting

It's important to note that the multi-level intent representation heads are pre-trained on limited labeled data. To fully leverage the valuable information embedded within unlabeled data, we employ pseudo-labeling techniques to iteratively select unlabeled samples as weak supervised signals for subsequent contrastive learning processes.

In the context of semi-supervised contrastive learning (4.3.2), reliable pseudo-labeled data are amalgamated with labeled data, augmenting the pool of potential contrastive samples. Regarding prototypical contrastive learning (4.3.3), pseudo-labels are employed to enrich the learning process of the prototype matrix. Furthermore, the integration of pseudo-labeled data introduces supplementary constraints to mitigate overfitting and enhance the model's generalization performance to novel unseen categories.

However, it's important to note that the quality of pseudo-labels is crucial, as noisy or incorrect pseudo-labels can degrade model performance. It is crucial to ensure the accuracy and reliability of pseudo-labels to maintain the effectiveness of the model's training. We treat pseudo-labels with probabilities greater than a confidence threshold as true labels and use them as supervised signals to guide model training:

$$p(k|\mathbf{x}) > \sigma \tag{4}$$

where $p(k|\mathbf{x})$ denotes the probability that x belongs to class k, σ represents confidence probability. When a certain probability is greater than the confidence threshold, we have enough confidence to consider it as belonging to this category. The choice of confidence probability will impact the strength of the supervised signal and the introduced noise. A high confidence probability will result in inadequate supervised information, whereas a low confidence probability will introduce erroneous pseudo-labels. To simplify the threshold selection, a suitable value is chosen directly as $\sigma = 1 - 10^{-k}$. In this paper, we set $k = 2$. We also compare performance at different confidence thresholds.

4.3.2 Semi-supervised Contrastive Learning

During the training stage, we utilize distinct contrastive learning strategies for IND labeled data and OOD unlabeled data. Firstly, we compare confidence probabilities of pseudo-labels with a predefined confidence threshold, and if they are greater than this threshold, we consider them as reliable pseudo-labels. Unlabeled samples with reliable pseudo-labels are considered as labeled data. SCL is applied on the instance-level contrastive learning head f for labeled data, while unsupervised instance-level contrastive learning (ILCL) is performed on unlabeled data.

On the cluster-level contrastive learning head g, we perform cross-entropy loss (CE) for labeled data and perform cluster-level contrastive learning (CLCL) for OOD classes (OOD-setting) or

Dataset	Classes	Classes-IND	Classes-OOD	Training	Validation	Test
Banking	77	54	23	9003	1000	3080
Stackoverflow	20	14	6	12000	2000	6000

Table 1: Statistics of BANKING and STACKOVERFLOW datasets.

all classes(open-setting). During the training process, the number of unlabeled samples with reliable pseudo-labels will gradually increase.

4.3.3 Prototypical Contrastive Learning

Decoupling knowledge from different levels is beneficial for separating the features of the source domain and target domain, thereby improving the efficiency of transfer learning and reducing overfitting. However, previous work only disentangled the instance-level and cluster-level features and applied constraints on them independently, without considering the inherent connection between the two levels of features. Each sample potentially belongs to a cluster, and each cluster consists of a certain number of samples. In order to extract the relationship between instance features and cluster features and enhance the discrimination between clusters, we propose to maintain a cluster prototype matrix, which is of size $k * m$ and stores the prototype features of each cluster.

For each batch, the output of f is an $n * m$ matrix containing the feature vectors of each sample, n is the batch size and m is the feature vector dimension. The output of g is an $n * k$ matrix, with each row corresponding to the probability that a sentence belongs to each class, i.e., $p(k|x)$, and each column corresponds to the representation of a cluster. The cluster prototype matrix is computed by averaging the instance-level representations over all samples belonging to the class. For labeled data and unlabeled data with reliable pseudo-label, we use ground truth or pseudo-label, i.e.**Hard Label Constraint Feature Combination**; for other unlabeled data, we perform probability-weighted calculations, i.e.**Soft Semantic Weighted Feature Combination**. As shown in equation, G is the cluster-level feature matrix $[g_1; g_2; \cdots; g_m; 1_{y_{m+1}}, 1_{y_{m+2}}, \cdots, 1_{y_N}]$, F is the instance-level feature matrix $[f_1; f_2; \cdots; f_N]$, m is the number of labeled data in this batch. G' and F' are the cluster-level feature matrix and instance-level feature matrix of the augmented samples.

$$M_c = G^T F \tag{5}$$

The cluster prototype matrix M_c is a $k * m$ matrix consisting of the features of each clustering centre and can be written as $[m_1, m_2, \cdots, m_K]$. The obtained vector m_i is normalized and used as the clustering centre z_i as shown in equation.

$$z_c, i = \frac{m_i}{\|m_i\|_2} \tag{6}$$

After explicitly decoupling the cluster prototype vector z_i, the augmented features of each cluster prototype are used as positive samples, and the rest of the features are used as negative samples for contrastive learning at the cluster prototype level, as shown in the equation. Optimization of prototype contrastive loss (PCL) enables pulling apart different clusters and thus enhancing the discrimination between categories.

$$\ell_{i,j}^{pcl} = -log \frac{\exp\left(sim(z_{c,i}, z_{c,j})/\tau\right)}{\sum_{k=1}^{2N} 1_{i \neq k} \exp\left(sim(z_{c,i}, z_{c,k})/\tau\right)} \tag{7}$$

The final loss in the training process is obtained by combining SCL, ILCL, CE, CLCL and PCL.[2]

5 Experiments

5.1 Datasets

In order to fairly compare the effectiveness of the models, we use two public datasets STACKOVERFLOW(Xu et al., 2015) and BANKING(Casanueva et al., 2020). BANKING and STACKOVERFLOW are both single-domain intent datasets. BANKING consists of 13,083 queries covering 77 intents in the banking domain, while the STACKOVERFLOW dataset contains 20 intents related to the programming domain. Detailed statistics are shown in Table 1, the division of the training set, validation set and test set remains consistent with previous works. We take 30% categories as unknown categories in both datasets, and all data with known intent is labeled.

5.2 Baselines

We selected a series of semi-supervised methods as benchmarks for comparing OOD intent discov-

[2]We simply set the weight coefficients of each loss to 1. We compared the effects of different weights for supervised losses in appendix A.1.

Method	Banking-OOD			Stackoverflow-OOD			Banking-Open			Stackoverflow-Open		
	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI
DTC_BERT(Hsu et al., 2017)	45.76	42.88	74.97	57.83	32.31	37.29	42.56	31.72	69.12	52.7	35.19	49.3
KCL_BERT(Han et al., 2019)	47.61	36.5	64.51	41.33	28.74	34.42	64.87	54.52	80.07	63.43	50.42	61.83
MCL_BERT(Hsu et al., 2019)	45.87	34.85	62.83	42.39	27.04	33.71	65.39	55.21	79.53	63.55	47.51	57.18
CDAC+(Lin et al., 2020)	59.78	44.58	69.19	61.56	28.22	52.76	45.00	33.10	69.49	67.05	48.66	66.03
DeepAligned (Zhang et al., 2021b)	63.86	<u>52.84</u>	**73.66**	79.68	63.18	65.52	<u>74.84</u>	<u>64.37</u>	<u>84.86</u>	<u>76.77</u>	<u>59.42</u>	<u>71.97</u>
DKT(Mou et al., 2022b)	<u>66.50</u>	52.07	<u>72.22</u>	<u>82.22</u>	61.53	**67.05**	70.38	61.16	83.71	72.57	58.6	69.12
Llama2(Touvron et al., 2023)	27.82	45.26	3.25	71.24	<u>67.62</u>	48.63	25.13	43.21	2.06	69.26	66.00	40.64
PLPCL	**68.37**	**53.19**	72.04	**86.28**	**69.64**	<u>66.95</u>	**76.50**	**67.13**	**85.99**	**77.63**	**63.58**	**72.2**

Table 2: The results on two datasets and two task settings. Overall $1^{st}/2^{nd}$ in **bold**/<u>underline</u>.We randomly sample 30% of all classes as OOD intents for both datasets. Results are averaged over three random runs. ($p < 0.05$ under t-test)

ery and open intent discovery. We hope to perform lightweight operations and reduce reliance on external data, and MTP-CLNN (Zhang et al., 2022), which used a lot of externally labeled data during the pre-training phase, was not included in the comparison. For BANKING-OOD, results of CDAC+, DeepAligned and DKT are extracted from (Mou et al., 2022b), and others are obtained from the text open intent recognition platform (Zhang et al., 2021a). For BANKING-all and STACKOVERFLOW-all, all baselines use the same BERT backbone and the results of baselines except DKT are obtained from (Zhang et al., 2021a).

5.3 Evaluation Matrics

We use three cluster evaluation metrics ACC, ARI, and NMI to evaluate the model effect, followed by DeepAligned(Zhang et al., 2021b). To obtain the results of ACC, we use the Hungarian algorithm to map prediction categories to ground-truth.

5.4 Implementation Details

We use the pre-trained bert-base-uncased model as the backbone consistent with the previous work, and pooling the context embeddings for each token using GRU and dense layers. The feature vector dimension is 768, the dropout probability is 0.1, and the GRU layer number is 1. In OOD discovery, the batch size of IND pre-training is 128, the batch size of STACKOVERFLOW-OOD and BANKING-OOD in the OOD clustering stage are both 400. For open intent discovery, the batch_size is 128 for both STACKOVERFLOW-ALL and BANKING-ALL for the pre-training and training stages. As with DKT, the learning rate of the pre-training process is set to 5e-5 of the training process is set to 0.0003, and the instance-level feature dimension is 128. Therefore, the cluster prototype feature dimension is also 128. The training epochs for both pre-training and training stages are 100. The experiment was conducted on an RTX 2080 GPU, and the running process takes 4 hours.

5.5 Main Results

Table 2 shows the performance of different models under the two task settings of two datasets. Our model achieved the best performance in OOD clustering of both datasets, outperforms DKT by 1.87% (ACC) and 1.02% (ARI) in BANKING, 4.06% (ACC), 8.11% (ARI) in STACKOVERFLOW.

For open-settings, the performance of our model is significantly better than DKT. Our model outperforms DKT dy 5.06% (ACC), 4.98% (ARI), and 3.08% (NMI) on STACKOVERFLOW. And the results exceeded DKT by 6.12% (ACC), 5.97% (ARI) and 2.28% (NMI) on BANKING. This indicates that our improvement has a significant effect on all-class classification, enhancing the contrastive learning framework's ability to distinguish between multiple categories and compensating for the limitations caused by an insufficient contrastive sample. On the BANKING dataset, our method outperforms the best baseline by 1.66% (ACC), 2.76% (ARI), and 1.13% (NMI). On the STACKOVERFLOW dataset, our method outperforms the best baseline by 0.86% (ACC), 4.16% (ARI), and 0.23% (NMI). Our method outperformed previous approaches on both the banking and Stackoverflow datasets, indicating that it adapts well to single-domain intent classification datasets and has better discriminability for professional intents with semantically similar meanings.

5.6 Comparison with Large Language Model

Considering the rapid development of large language models in recent years, many tasks in natural language processing can be easily addressed by leveraging the text generation capabilities of these models. In the last row of the experiment results table 2, we compared our results with the latest local llama2-13B model(Touvron et al., 2023). Taking into account the input tokens' limitations and their relatively weaker clustering abilities, we employed the large models for classification tasks with the provision of category names.

The example of a prompt template is shown in the appendix A.3 The results indicate that the performance of LLM is inferior to our method in both settings of the two datasets. On the STACK-OVERFLOW dataset with few categories, LLM outperforms some previous methods. However, on the banking dataset with a larger number of categories, LLM clustering shows poor clustering performance.

5.7 Ablation study and further analysis

Table 3 shows the effects of each module in our model, experimenting on BANKING-ALL. The results show that including SCL and CE during training helps to fully utilize the supervised signal. The absence of the supervised signal will result in a partial loss of pre-training information and a significant decrease in effectiveness. The addition of contrastive learning of prototypes (PCL) improved the model's performance by 3.91% (ACC), 4.74% (ARI), and 1.65% (NMI), indicating that explicitly decoupling and separating the cluster centers is beneficial for distinguishing and separating different category features in the feature space, increasing the distance between clusters, reducing the probability of confusion, and enhancing clustering performance. The results of the following cluster visualization further demonstrate this point. The addition of confident pseudo-labels (PL) improved the model's performance by 2.21% (ACC), 1.23% (ARI), and 0.63% (NMI), indicating that gradually including samples with sufficiently high confidence in the supervised signal during model iteration is beneficial for obtaining prior information, compensating for the limitations of simple sample scattering in unsupervised contrastive learning.

	ACC	ARI	NMI
ILCL,CLCL	46.07	36.79	37.00
+SCL,CE	70.38	61.16	83.71
+PCL	74.29	65.90	85.36
+PL	**76.50**	**67.13**	**85.99**

Table 3: Ablation Study of different training objectives.

Table 4 shows the effect of different confidence thresholds on the effectiveness of the model. When the threshold is 1, no pseudo-labels are used. When

σ	ACC	ARI	NMI
1	74.29	65.90	85.36
0.99	**76.50**	**67.13**	85.99
0.9	76.43	66.25	**86.16**
0.8	74.38	65.25	85.48
0.7	73.18	64.75	85.35
0.5	70.78	62.39	84.19
0	66.17	17.66	80.72

Table 4: Results under Different Confidence Thresholds.

the threshold is 0, pseudo-labels are used for all samples.

Figure 3 illustrates the clustering performance (ACC) of various models at different labeled ratios for IND intents while maintaining an IND category ratio of 70%. The results demonstrate the robust performance of our models across different labeling ratios. More results showing the effects of labeled ratio and known cluster ratio can be checked in appendix A.1.

Figure 4 demonstrates the performance of our model on the banking dataset under different weights of supervised contrastive loss, showing that our model is insensitive to loss weights.

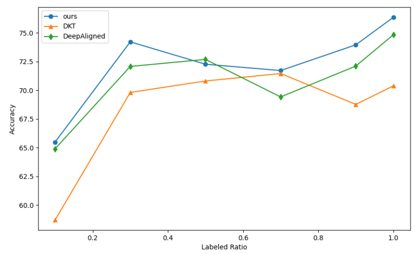


Figure 3: Influence of the labeled ratio on BANKING dataset.

6 Conclusion

In this paper, we have proposed a pseudo-label enhanced prototypical contrastive learning approach for both open intent discovery and OOD intent discovery. The pseudo-label filtering strategy enhances supervised signal during training process, while the prototypical contrastive learning module addresses the isolation issue between two independent contrastive learning heads. Compared with previous methods, our approach provides better knowledge transfer. Experiments on two task settings and two benchmark datasets demonstrate the effectiveness of our proposed method. We hope to explore more self-supervised methods for OOD and open intent discovery in the future.

Limitations

Our model considers the inherent connection between two levels of features while decoupling instance-level and cluster-level features. We also consider the potential relation between different unlabeled samples. Our proposed model is effective in both open intent discovery and OOD intent discovery scenarios. In order to determine the update of categories, it is necessary to accurately predict the number of categories. Although we used intent understanding datasets in our work, our approach also holds promise as a novel contrastive learning method that can be applied widely to clustering scenarios, such as topic classification and multi-view clustering.

References

Wenbin An, Feng Tian, Qinghua Zheng, Wei Ding, QianYing Wang, and Ping Chen. 2023. Generalized category discovery with decoupled prototypical network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 12527–12535.

Iñigo Casanueva, Tadas Temčinas, Daniela Gerz, Matthew Henderson, and Ivan Vulić. 2020. Efficient intent detection with dual sentence encoders. *arXiv preprint arXiv:2003.04807*.

Dilek Hakkani-Tür, Yun-Cheng Ju, Geoffrey Zweig, and Gokhan Tur. 2015. Clustering novel intents in a conversational interaction system with semantic parsing. In *Sixteenth Annual Conference of the International Speech Communication Association*.

Mark Hamilton, Zhoutong Zhang, Bharath Hariharan, Noah Snavely, and William T Freeman. 2022. Unsupervised semantic segmentation by distilling feature correspondences. *arXiv preprint arXiv:2203.08414*.

Kai Han, Andrea Vedaldi, and Andrew Zisserman. 2019. Learning to discover novel visual categories via deep transfer clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8401–8409.

Yen-Chang Hsu, Zhaoyang Lv, and Zsolt Kira. 2017. Learning to cluster in order to transfer across domains and tasks. *arXiv preprint arXiv:1711.10125*.

Yen-Chang Hsu, Zhaoyang Lv, Joel Schlosser, Phillip Odom, and Zsolt Kira. 2019. Multi-class classification without multi-class labels. *arXiv preprint arXiv:1901.00544*.

Tianyang Hu, Zhili Liu, Fengwei Zhou, Wenjia Wang, and Weiran Huang. 2022. Your contrastive learning is secretly doing stochastic neighbor embedding. *arXiv preprint arXiv:2205.14814*.

Zhizhong Huang, Jie Chen, Junping Zhang, and Hongming Shan. 2022. Learning representation for clustering via prototype scattering and positive sampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Rajat Kumar, Mayur Patidar, Vaibhav Varshney, Lovekesh Vig, and Gautam Shroff. 2022. Intent detection and discovery from user logs via deep semi-supervised contrastive clustering. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1836–1853.

Yunfan Li, Peng Hu, Zitao Liu, Dezhong Peng, Joey Tianyi Zhou, and Xi Peng. 2021. Contrastive clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 8547–8555.

Ting-En Lin, Hua Xu, and Hanlei Zhang. 2020. Discovering new intents via constrained deep adaptive clustering with cluster refinement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8360–8367.

Pengfei Liu, Youzhang Ning, King Keung Wu, Kun Li, and Helen Meng. 2021. Open intent discovery through unsupervised semantic clustering and dependency parsing. *arXiv preprint arXiv:2104.12114*.

Qingkai Min, Libo Qin, Zhiyang Teng, Xiao Liu, and Yue Zhang. 2020. Dialogue state induction using neural latent variable models. *arXiv preprint arXiv:2008.05666*.

Yutao Mou, Keqing He, Yanan Wu, Pei Wang, Jingang Wang, Wei Wu, Yi Huang, Junlan Feng, and Weiran Xu. 2022a. Generalized intent discovery: Learning from open world dialogue system. *arXiv preprint arXiv:2209.06030*.

Yutao Mou, Keqing He, Yanan Wu, Zhiyuan Zeng, Hong Xu, Huixing Jiang, Wei Wu, and Weiran Xu. 2022b. Disentangled knowledge transfer for ood intent discovery with unified contrastive learning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 46–53.

Yutao Mou et al. 2022c. Watch the neighbors: A unified k-nearest neighbor contrastive learning framework for ood intent discovery. *arXiv preprint arXiv:2210.08909*.

Xiang Shen, Yinge Sun, Yao Zhang, and Mani Najmabadi. 2021. Semi-supervised intent discovery with contrastive learning. In *Proceedings of the 3rd Workshop on Natural Language Processing for Conversational AI*, pages 120–129.

Chen Shi, Qi Chen, Lei Sha, Sujian Li, Xu Sun, Houfeng Wang, and Lintao Zhang. 2018. Autodialabel: Labeling dialogue data with unsupervised learning. In *Proceedings of the 2018 conference on empirical methods in natural language processing*, pages 684–689.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.

Nikhita Vedula, Nedim Lipka, Pranav Maneriker, and Srinivasan Parthasarathy. 2019. Towards open intent discovery for conversational text. *arXiv preprint arXiv:1904.08524*.

Jiaming Xu, Peng Wang, Guanhua Tian, Bo Xu, Jun Zhao, Fangyuan Wang, and Hongwei Hao. 2015. Short text clustering via convolutional neural networks. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pages 62–69.

Hanlei Zhang, Xiaoteng Li, Hua Xu, Panpan Zhang, Kang Zhao, and Kai Gao. 2021a. Textoir: An integrated and visualized platform for text open intent recognition. *arXiv preprint arXiv:2110.15063*.

Hanlei Zhang, Hua Xu, Ting-En Lin, and Rui Lyu. 2021b. Discovering new intents with deep aligned clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14365–14373.

Yuwei Zhang, Haode Zhang, Li-Ming Zhan, Xiao-Ming Wu, and Albert Lam. 2022. New intent discovery with pre-training and contrastive learning. *arXiv preprint arXiv:2205.12914*.

A Appendix

A.1 Further Analysis of loss weights and settings

Figure 4 demonstrates the performance of our model on the banking dataset under different weights of supervised contrastive loss, showing that our model is insensitive to loss weights. Figure 7 illustrates the impact of different labeled ratios and known cluster ratios on the model performance.

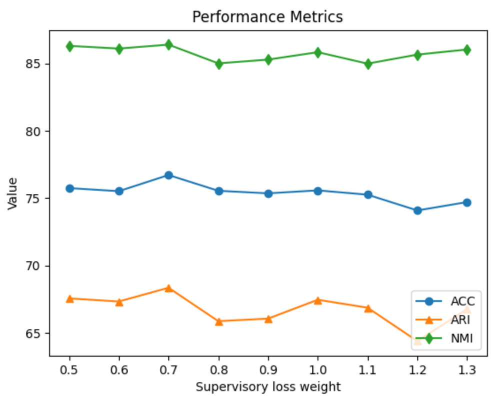


Figure 4: Influence of the supervisory loss weight on BANKING dataset

A.2 Visualization

Figure 5 shows the clustering visualization results of DKT and our model on BANKING-ALL and BANKING-OOD. For fair comparison, we use the same representation after the pooling layer. We can find that after adding contrastive learning for prototype and reliable pseudo label, while keeping the samples of the same cluster compact, the distance between different clusters is widened, and the different clusters become scattered on the whole feature space.Unlike scattering of unlabeled samples, the premise for contrastive learning in cluster prototype is that each cluster has its own unique features, and cluster center scattering aims to separate these features.

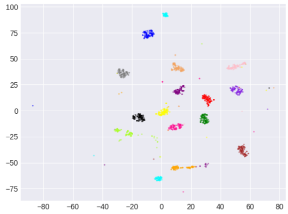
Figure 6 shows the visualization results of the confusion matrix for DKT and our model on BANKING-ALL, with a total of 77 categories in the test set, and we show the first 20 categories. It can be found that the DKT model may completely confuse two certain categories, i.e. the samples of two certain categories are grouped into the same cluster. However, our model avoids this problem well, and rarely there is no correct sample in a certain category.

A.3 Prompt Template of LLM

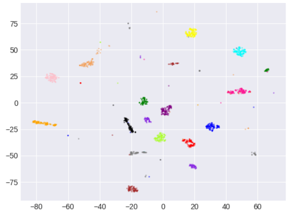
The prompt template is shown in the table5.

Below is an instruction that describe a task. Write a response that appropriately completes the request. ###instruction: Please give the intent label for the following sentences. Select one label in the set {...} For example: Input: Can I exchange currencies with this app? . Output: { intent_label:"exchange_via_app"} ###question: Input: {data sample} Provide intent label in JSON format with the following keys: intent_label ###Response:

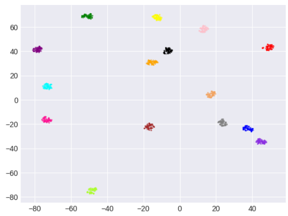
Table 5: An example of the prompt templates we used.



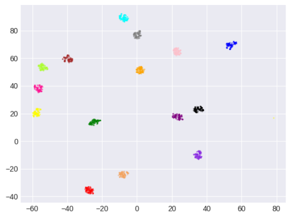
(a) DKT_OOD



(b) ours_OOD

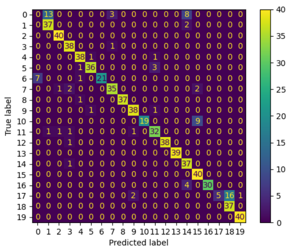


(c) DKT_ALL

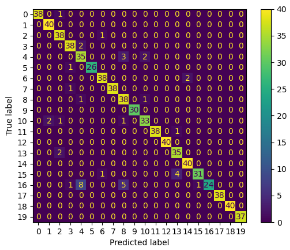


(d) ours_ALL

Figure 5: OOD intent visualization of different models. We use the same test set of Banking-ALL.



(a) DKT_cm



(b) ours_cm

Figure 6: Confusion matrix visualization of different models.We use the same test set of Banking-ALL.

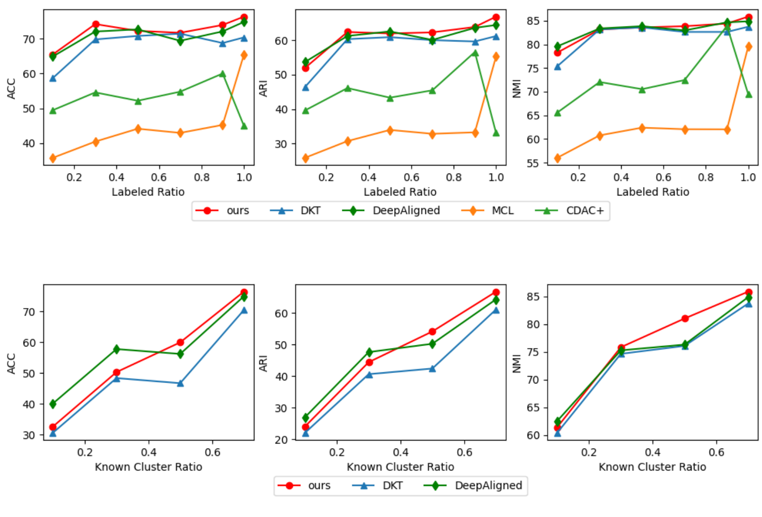


Figure 7: Influence of the labeled ratio and known cluster ratio on BANKING dataset.