LA-UCL: LLM-Augmented Unsupervised Contrastive Learning Framework for Few-Shot Text Classification

Jing Zhang*, Hui Gao*, Peng Zhang† , Boda Feng, Wenmin Deng, Yuexian Hou
College of Intelligence and Computing, Tianjin University, Tianjin, China
{zhang_jing, pzhang}@tju.edu.cn

Abstract

The few-shot tasks require the model to have the ability to generalize from a few samples. However, due to the lack of cognitive ability, the current works cannot fully utilize limited samples to expand the sample space and still suffer from overfitting issues. To address the problems, we propose a **L**LM-**A**ugmented **U**nsupervised **C**ontrastive **L**earning Framework (**LA-UCL**), which introduces a cognition-enabled Large Language Model (LLM) for efficient data augmentation, and presents corresponding contrastive learning strategies. Specifically, in the self-augmented contrastive learning module, we construct a retrieval-based in-context prompt scheme by retrieving similar but different category data from the original samples, guiding the LLM to generate more discriminative augmented data. Then, by designing group-level contrastive loss to enhance the model's discriminative ability. In the external-augmented contrastive learning module, we utilize web knowledge retrieval to expand the sample space and leverage LLM to generate more diverse data, and introduce sample-level contrastive loss for unlabeled data to improve the model's generalization. Experimental results on six datasets show that our model exceeds the baseline models.

Keywords: Data augmentation, Contrastive learning, Few-shot learning

1. Introduction

The few-shot learning in resource constraint settings, which gets rid of the limitation of data labeling cost for the deep learning (Wang et al., 2020; Murty et al., 2021), has attracted a lot of attention. Compared with traditional text classification tasks, few-shot text classification requires models to learn new conceptual categories quickly and efficiently with few examples, just like humans, which poses a greater challenge to the cognitive ability and generalization ability of models (Lake et al., 2019; Cao et al., 2021).

Early studies on few-shot learning mostly used meta-learning framework (Bao et al., 2020). Yu et al. (2018) proposed an adaptive metric learning approach that automatically determines the best-weighted combination from a set of metrics obtained from meta-training tasks. Geng et al. (2019) proposed a novel Induction Network to learn such a generalized class-wise representation, by innovatively leveraging the dynamic routing algorithm in meta-learning. In recent years, the contrastive learning framework has gradually shown its advantages. Chen et al. (2022) proposed the contrastive learning framework significantly improves the text discrimination ability compared with meta-learning, which proves that data augmentation combined with contrastive learning is effective in dealing with few-shot tasks.

However, due to the lack of cognitive ability and prior knowledge of generative data augmentation, existing generative data augmentation algorithms

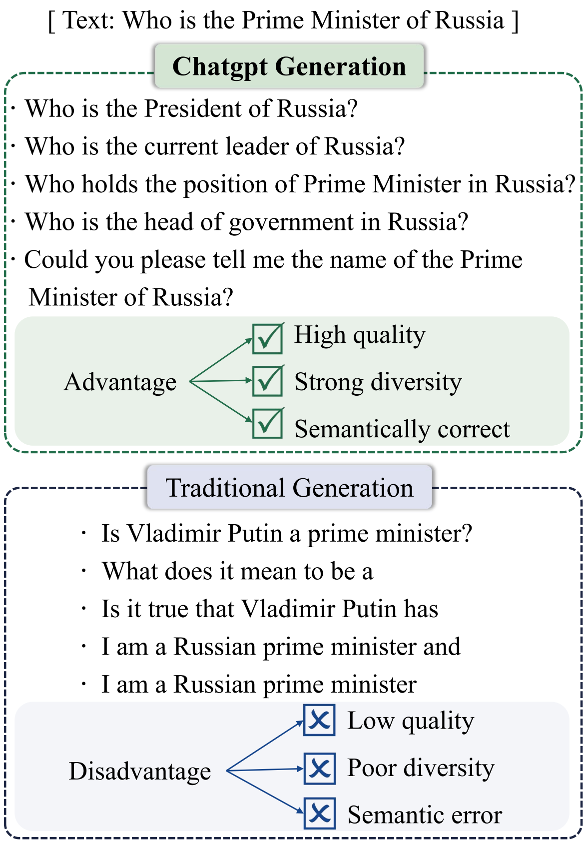


Figure 1: Augmented samples generated by traditional models and ChatGPT.

struggle to generate diverse and high-quality data, resulting in: 1) **Poor discrimination ability**: In the data augmentation stage, the traditional generation model is easy to generate new samples that are roughly the same as the old sample (as shown in

*These authors contributed equally.
† Corresponding author: Peng Zhang

LREC-COLING 2024, pages 10198–10207

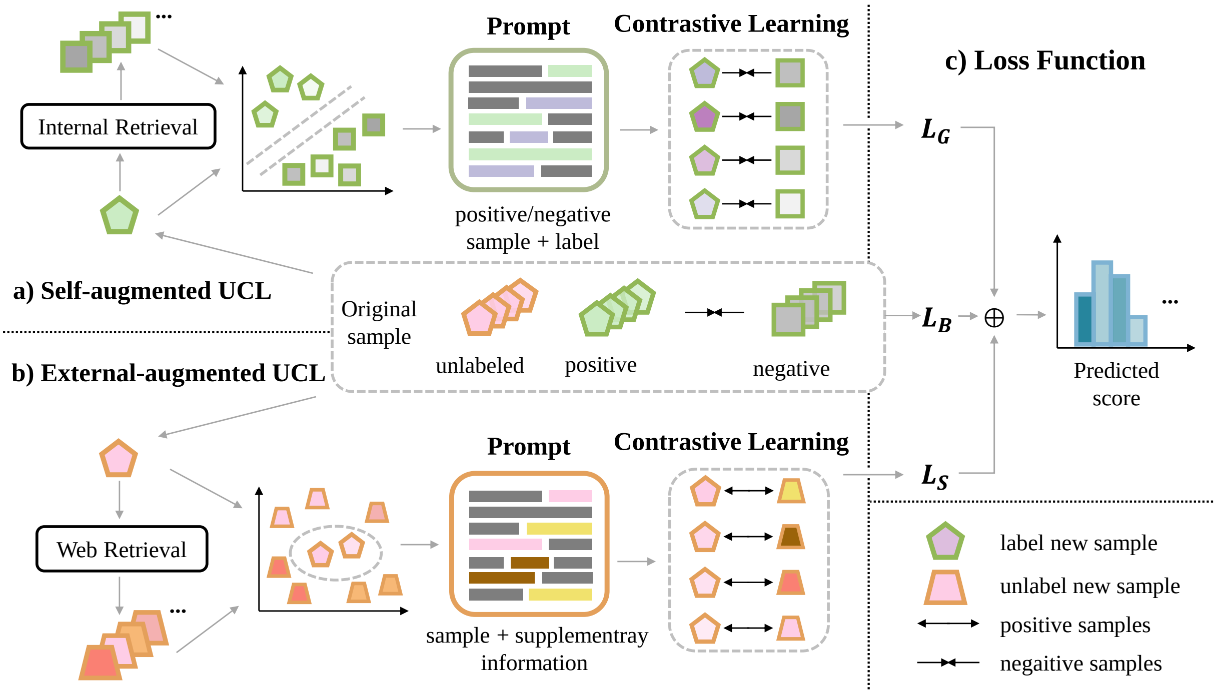


Figure 2: The overall of LA-UCL.

Figure 1), resulting in the model failing to capture diversified sample features under the contrastive learning. 2) **Training overfitting**: A single and fuzzy sample generation result will aggravate the overfitting problem of the model in the few-shot scenario, affecting the feature learning of samples in the subsequent contrastive learning framework.

To solve the challenges, we incorporate high-capacity Large Language Models (LLMs) known for their cognitive abilities into a few-shot contrastive learning framework for data augmentation, to overcome the problems related to category distinction and overfitting in contrastive learning methods. Specifically, we separately propose two data enhancement algorithms based on generative-style prompts, which are integrated into an unsupervised contrastive learning (UCL) framework to optimize loss function:

1) **Self-augmented UCL**: In order to make the LLM generate more discriminating samples, we use the original instance as a query to retrieve several different types of instances with the highest query-related degree in the dataset. This approach effectively expands the sample space, avoids the illusion of LLM, and helps the generated data representation be more discernable; Based on self-augmented contrastive learning, we design a group-level contrastive learning method, which proposes a contrastive learning loss based on the current batch group and base class groups, takes the original data as the support set, and uses LLM augmented data as the query set, which improves the model's ability to distinguish the base classes.

2) **External-augmented UCL**: To combat overfitting, we designed sample-level contrastive learning loss based on unlabeled samples and LLM-augmented data. We use web retrieval to find out the external knowledge information related to the unlabeled instances and construct in-context prompts to guide the LLM to generate more diversified and accurate interpretations. This approach alleviates the overfitting issue faced by contrastive learning. Additionally, UCL framework is added to the batch contrastive learning loss (Khosla et al., 2020), and learnable scalars are used to adaptively adjust the impact of three types of loss.

Our experiments surpass recent few-shot text classification models, incorporating contrastive learning and traditional learning methods. To summarize, our main contributions are as follows:

- We propose two retrieval-based in-context prompt to guide the cognition of LLM, thereby generating discriminative and diverse augmented data.
- We propose a generative data augmented unsupervised contrastive learning framework to improve the model's ability to discriminate diverse samples while alleviating overfitting problems.
- We conduct experiments on six text classification datasets and show that LA-UCL outperforms our baseline models.

2. Methodology

2.1. Problem Formulation

For a typical n-way k-shot text classification problem, there are three datasets: training set Y_{train}, verification set Y_{val} and testing set Y_{test}. One dataset contains a set of classes, and the class sets of the Y_{train}, Y_{val} and Y_{test} have no intersection. An episode of contrastive learning contains the support set $\mathcal{S}$ and the query set $\mathcal{Q}$. In the training, validation, or testing phase, we randomly select n classes from their respective class sets, where each class selects k labeled texts are selected to create the support set and m unlabeled texts are selected into the query set.

2.2. Self-augmented Unsupervised Contrastive Learning

In order to improve the discriminative ability of few-shot contrastive learning methods for similar texts, we propose a prompt scheme based on the Mixup (Guo et al., 2019) idea, which enables LLM to generate data-augmented samples, and in turn, proposes a unsupervised contrastive learning loss based on internal knowledge data augmentation.

2.2.1. LLM Data Augmentation with Mixup Strategy

The principle of Mixup idea is to combine positive sample features and negative sample features and get new samples, which can add robustness to the model and facilitate data diversity. We design the prompt of ChatGPT based on the Mixup idea. Specifically, for an instance s_i, the category is y_i, we retrieve the negative samples u_i, which have the highest similarity with it but are not in the category of y_i in the dataset by the bm25 method, the specific equation is shown as follows:

$$u_i = \{u_{i,1}, u_{i,2}, ..., u_{i,K} | y_{i,k} \neq y_i\} \quad (1)$$

Based on the above retrieval, we design a novel in-context prompt for ChatGPT to generate diverse positive samples, The style is shown as Prompt 2.2.1.

In Prompt 2.2.1, we need to be given three aspects of information about the prompt, namely labeled positive sample data s_i, labels y_i, K negative sample data $u_{i,k}$ and their corresponding labels $y_{i,k}$. The reason for adding labels is that for some classification tasks, the labels themselves also contain semantic information. In addition, we provide ChatGPT with hand-crafted cases as in-context to guide ChatGPT in understanding task requirements and standardizing output formats. The number of cases and the size of K are limited by the ChatGPT decoding length.

As shown in Figure 2, our prompt method is able to give the large model a control group, allowing the LLM to generate positive-sample augmented data that are more diverse but closer to the instances. Ultimately, based on LLM's meta-knowledge and generative capabilities, the designed prompts provided LLM with internal knowledge of the dataset, guiding the ChatGPT to generate K more diverse augmented data.

Prompt 2.2.1

You are a sample generator. For a given label and its positive sample and negative sample texts, please compare them and generate K novel positive samples based on your knowledge reserve. The examples are as follows.
Case 1: {
Label & Samples
}. Please generate K positive samples.
response:
1. ...
...
K. ...

The current instruction is as follows: {
Label: y_i
Positive sample: s_i
Negative sample of category $y_{i,1}$: $u_{i,1}$
...
Negative sample of category $y_{i,K}$: $u_{i,K}$
}. Please generate K positive samples.
response:

2.2.2. Group-Level Contrastive Loss

We utilize a unsupervised comparative learning loss of labelled samples in order to further improve the generalisation of contrastive learning and make its representation more discriminative.

In each episode, we utilize instances and augmented instances of base classes to form N_G groups $\{(\mathcal{S}_1, \mathcal{Q}_1), .., (\mathcal{S}_{N_G}, \mathcal{Q}_{N_G})\}$. Then, we use the support set of the training batch (episode) and its corresponding augmented samples form a group $(\mathcal{S}_{N_G+1}, \mathcal{Q}_{N_G+1})$. In order to further improve the model's ability to distinguish categories, we use unsupervised contrastive learning for $(N_G + 1)$ group interactions.

$$\mathcal{L}_G = - \sum_{g=1}^{2(N_G+1)} log \frac{e^{\frac{z_g \cdot z'_g}{\tau}}}{e^{\frac{z_g \cdot z'_g}{\tau}} + \sum\limits_{z_{g'} \neq z'_g} e^{\frac{z_g \cdot z_{g'}}{\tau}}} \quad (2)$$

where τ is a temperature factor that scales the inner products. Each group has two sets, so $g \in$

$[1, ..., 2(N_G + 1)]$, when $g \leq N_G + 1$, g represents the $\mathcal{S}$ set of the g-th group. When $g \geq N_G + 1$, g represents the $\mathcal{Q}$ set of the $(g - N_G - 1)$-th group. In addition, g' refers to the matching set of g-th set, that is, when g-th set is $\mathcal{Q}$, g' is the matching $\mathcal{S}$, when g-th set is $\mathcal{S}$, g' is the matching $\mathcal{Q}$. z_g represents the average representation of the g-th set, while z'_g represents the average representation of the g-th matching set. $z_{g'}$ represents average instance embeddings in $2(N_G+1)$ sets that are not equal to z'_g. The representation of each instance is the $[CLS]$ vector of the BERT model.

This unsupervised contrastive learning method aims to shorten the distance between training batch samples and augmented data, widen the distance from different base class representations, and enhance the model's discrimination ability by increasing the interaction with the base classes.

2.3. External-augmented Unsupervised Contrastive Learning

To avoid the overfitting problem during batch learning, we use external knowledge retrieval to construct cues for unlabelled text, which induces the LLM to generate more diverse data in a larger sample space and propose unsupervised diverse contrastive learning loss.

Prompt 2.3.1

You act as a paraphrase tool. Your role involves understanding the user's input and, using your knowledge, along with supplementary information retrieved from the web, generating K different variants of the user's input with the same meaning. The examples are as follows.
Case 1: {
Input & Supplementary information
}. Please generate K positive samples.
response:
1. ...
...
K. ...
The current instruction is as follows: {
Input: s_i
Supplementary information: w_i
}.
Please directly generate K positive samples.
response:

2.3.1. LLM Data augmentation with External Knowledge

We employ external knowledge retrieval to expand the information space of unlabeled instances, guiding the LLM in generating diverse data. Specifically, we use Bing web search to select the top-ranked knowledge entries as supplementary information w_i. After obtaining the supplementary information [optional] of instance s_i, we design a novel in-context prompt for ChatGPT.

In the Prompt 2.3, the input s_i of the user is an unlabeled instance. While large models possess exceptional knowledge reservoirs and cognitive capabilities, in order to avoid the pitfalls of model hallucination, we need to provide supplementary information w_i retrieved from the web to guide the LLM in making accurate interpretations, thereby enabling the generation of more diverse and accurate meanings for unlabeled instances.

As shown in Figure 2, our data augmentation method theoretically expands the information space of the instance and guides the understanding of large models. In conclusion, we leverage optional external knowledge and unlabeled instances to create prompts, enabling us to generate more diverse and enriched instances. This approach is instrumental in effectively mitigating overfitting issues in contrastive learning.

2.3.2. Sample-Level Contrastive Loss

To alleviate the overfitting issue, we adopt unlabeled samples and their LLM-augmented samples to construct an unsupervised contrastive learning loss.

The unsupervised contrastive learning method still involves support sets and query sets $(\mathcal{S}_S, \mathcal{Q}_S)$, where the support set consists of N_S unlabeled instances, and the query set contains the corresponding LLM-augmented instances. We employ the following contrastive learning approach to bring diverse paraphrase of the same instance closer while pushing other instances and their corresponding diverse representations farther apart, thereby mitigating the issue of overfitting (Chen et al., 2022).

$$\mathcal{L}_S = -\sum_{s=1}^{2N_S} log \frac{e^{\frac{z_s \cdot z'_s}{\tau}}}{e^{\frac{z_s \cdot z'_s}{\tau}} + \sum_{z_{s'} \neq z'_s} e^{\frac{z_s \cdot z_{s'}}{\tau}}} \tag{3}$$

where z_s is the text representation of s-th sample in $(\mathcal{S}_S, \mathcal{Q}_S)$, z'_s is text representation of the s-th matching sample and τ is the temperature.

2.4. Overall Loss Function

The overall loss consists of unsupervised contrastive learning and supervised contrastive learning. Based on the labeled support set and a query set $(\mathcal{Q}, \mathcal{S})$ within the current training batch, we applied a batch supervised contrastive learning method (Khosla et al., 2020):

$$\mathcal{L}_B = -\sum_{b}^{N_B} \frac{1}{\frac{N_B}{n}-1} log \frac{\sum\limits_{y_b=y_t} e^{\frac{z_b \cdot z_t}{\tau}}}{\sum\limits_{y_b=y_t} e^{\frac{z_b \cdot z_t}{\tau}} + \sum\limits_{y_b \neq y_{t'}} e^{\frac{z_b \cdot z_{t'}}{\tau}}} \tag{4}$$

where N_B is the sum of the number of support set instances and the number of query set instances in an episode (batch) and n is the number of classes in episode. This method can bring text representations of the same class closer and make text representations of different classes farther away. Therefore, the overall loss can be defined as:

$$\mathcal{L} = \mathcal{L}_B + \alpha \mathcal{L}_G + \beta \mathcal{L}_S \tag{5}$$

where α and β are trainable scalars. Through the method of combining contrastive learning and LLM knowledge enhancement, the model's discriminability of difficult-to-distinguish samples is effectively improved, and the over-fitting problem of the model is also alleviated.

Moreover, about obtaining the predicted label of a instance q in query set, we calculated the dot product similarity between the query representation $\mathrm{BERT}(q)$ and all instance embeddings in the support set. The label of the support instance with the highest similarity is the predicted label of q.

3. Experiment

3.1. Datasets

We conducted experiments on six text classification datasets:

Banking77 (Casanueva et al., 2020) is a fine-grained intent classification dataset designed for a specific banking domain. It consists of 13,083 user utterances divided into 77 distinct intents.

HWU64 (Liu et al., 2021) is a across multi-domain intent classification dataset, which contains 11, 036 examples for 64 intents in 21 domains.

Clinic150 (Chen et al., 2022) intent classification dataset spans 150 intents and 23,700 examples across 10 domains.

Liu57 (Liu et al., 2021) is a multi-domain intent classification dataset collected on Amazon Mechanical Turk, which is composed of 25, 478 user utterances over 54 classes and 64 intents.

HuffPost (Wu et al., 2020) is a collection of news headlines published on HuffPost between 2012 and 2018, which consists 36, of 900 news with 41 classes.

Reuters (Wu et al., 2020) dataset is collected from shorter Reuters articles from 1987, which consist of 31 classes and 620 news.

3.2. Baseline Models

We compare the LA-UCL with following baselines:

Prototypical Networks (Snell et al., 2017) is a metric-based meta-learning method for few-shot classification that aims to align query instances with class prototypes.

PROTAUGMENT (Dopierre et al., 2021) employs a short-text paraphrasing model to generate augmented data and incorporates an instance-level unsupervised loss into the prototypical networks. It has two variants, i.e., PRO (unigram/bigram).

ContrastNet (Chen et al., 2022) is a few-shot text classification framework that learns discriminative text representations via contrastive learning.

MAML (Finn et al., 2017) facilitates fast adaptation of deep neural networks to new tasks with limited data based on meta-learning.

Induction Networks (Geng et al., 2019) introduces dynamic routing algorithm to learn the class-level representation.

HATT (Gao et al., 2019) combines attention mechanisms and prototypical networks to address few-shot relation classification.

DS-FSL (Bao et al., 2020) aims to enhance the transferability of features by mapping distribution signatures to attention scores.

MLADA (Han et al., 2021) adopts meta-learning for quick adaptation and domain adversarial training for domain alignment.

3.3. Experimental Settings

We evaluate LA-UCL on 5-way 1-shot and 5-way 5-shot text classification settings. Reference the work of Chen et al. (2022), for the intent classification datasets, we report the average accuracy over 600 samples sampled from the test set. For the news classification datasets, we report the average accuracy of over 1000 samples sampled from the test set. Each experimental setup was run five times using the re-split datasets.

All experiments are run on NVIDIA Tesla V100 PCIe 32GB GPUs, and we leverage Pytorch framework to implement the proposed models. For training, we use Adam (Kingma and Ba) to optimize the proposed model with an initialized learning rate of 1e-6. On the 4 intent classification datasets, we use their respective pre-trained BERT-based language model provided in (Devlin et al., 2018) as the encoders for text representation. For the news classification datasets, we use the pure pre-trained bert-base-uncased model as the encoder for text representation. For each episode during training, we randomly sample $N_G = 10$ groups and $N_S = 10$ unlabeled texts to calculate the group-level contrastive regularization loss and sample-level contrastive regularization loss. The temperature factors t_B,t_G and t_S of loss L_B, L_G and L_S are set

Table 1: The few-shot text classification results on the Banking77, HWU64, Liu and Clinic150 datasets.

Dataset	**Banking77**		**HWU64**		**Liu57**		**Clinic150**		**AVerage**	
Models	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Prototypical Net	86.28	93.94	77.09	89.02	82.76	91.37	96.05	98.61	85.55±2.20	93.24±1.22
PROTAUGMENT	86.94	94.5 0	82.35	91.68	84.42	92.62	94.85	98.41	87.14±1.36	94.30±0.60
PRO (bigram)	88.14	94.70	84.05	92.14	85.29	93.23	95.77	98.50	88.30±1.43	94.64±0.59
PRO (unigram)	89.56	94.71	84.34	92.55	<u>86.11</u>	93.70	96.49	**98.74**	89.13±1.13	94.92±0.57
ContrastNet	<u>91.18</u>	<u>96.40</u>	<u>86.56</u>	<u>92.57</u>	85.89	<u>93.72</u>	<u>96.59</u>	98.46	<u>90.06±1.02</u>	<u>95.29±0.53</u>
LA-UCL	**92.63**	**97.25**	**89.46**	**94.04**	**87.49**	**94.34**	**97.03**	<u>98.73</u>	**91.65±1.04**	**96.09± 0.62**

Table 2: The few-shot text classification results on the HuffPost and Reuters datasets.

Dataset	**HuffPost**		**Reuters**	
Models	1-shot	5-shot	1-shot	5-shot
MAML	35.9	49.3	54.6	62.9
PrototypicalNet	35.7	41.3	59.6	66.9
InductionNet	38.7	49.1	59.4	67.9
HATT	41.1	56.3	66.0	43.2
DS-FSL	43.0	63.5	81.8	96.0
MLADA	45.0	64.9	82.3	96.7
ContrastNet	<u>53.06</u>	<u>65.32</u>	<u>86.42</u>	<u>95.33</u>
LA-UCL	**54.94**	**68.96**	**87.70**	**96.61**

Table 3: Ablation experiments on Liu and Reuters datasets

Dataset	**Liu57**		**Reuters**	
Models	1-shot	5-shot	1-shot	5-shot
w / o $\mathcal{L}_G$& $\mathcal{L}_S$	81.39	92.87	82.44	92.81
w / o $\mathcal{L}_G$	82.02	93.15	83.20	94.95
w / o LLM	86.14	93.47	85.54	95.32
w / o Retrieval	86.90	93.91	85.92	95.54
LA-UCL	**87.49**	**94.34**	**87.70**	**96.61**

from 3.0 to 8.0, respectively. Moreover, since the intent data set is relatively subjective, we did not use the web to retrieve information in the data augmentation method of external-augmented UCL.

3.4. Main Result

As presented in Table 1 and Table 2, this study presents comparative experimental findings on four intention datasets and two news datasets using both the 5-way 1-shot and 5-way 5-shot settings. The results of baselines comprise the original data reported in the original paper. The table highlights the optimal results within each column with **bold** numbers, while the <u>underlined</u> numbers indicate sub-optimal results. In addition, we report the average results and average standard deviation in the intent task.

The experimental results of the intention classification task demonstrate the superior performance of the proposed LA-UCL model over all baselines, including meta-learning and contrastive learning approaches. When compared to meta-learning frameworks like the PRO(unigram) model, our approach achieved an average improvement of 2.8% in the 1-shot setup and 1.2% in the 5-shot setup. Compare with ContrastNet, LA-UCL enhances the 1-shot setting by 1.77% through the introduction of retrieval in-context prompts and an unsupervised contrastive learning loss that fosters interaction among base classes. Specifically on the HWU64 dataset, LA-UCL achieves improvements of 3.35% in the 5-way 1-shot scenario and 1.59% in the 5-way 5-shot scenario. These experimental findings provide strong evidence for the effectiveness of LA-UCL in tackling few-shot learning challenges in short text classification tasks.

For news tasks, our results are better than all baselines. In addition, it is worth noting that we use the preprocessed data set provided by ContrasNet, so we do not utilize models that preprocess data by themselves as baselines. In comparison to MLADA and other meta-learning algorithms, LA-UCL exhibits significant improvements in the 5-way 1-shot setting for both the HuffPost and Reuters datasets, with enhancements of 22.09% and 6.3%, respectively. Similarly, when compared to the ContraNet, our model achieves improvements of 3.54% and 1.48%, respectively. Furthermore, in the context of long text classification tasks like Reuters, LA-UCL effectively enhances the performance in modeling few-shot problems.

3.5. Ablation Study

To verify the effect of the model improvement mechanism, we conducted an ablation experiment, as shown in Table 3. 'w/o $\mathcal{L}_G\&\mathcal{L}_S$' means that two unsupervised contrastive learning losses are eliminated, 'w/o $\mathcal{L}_G$' represents the removal of self-augmented unsupervised contrastive learning loss

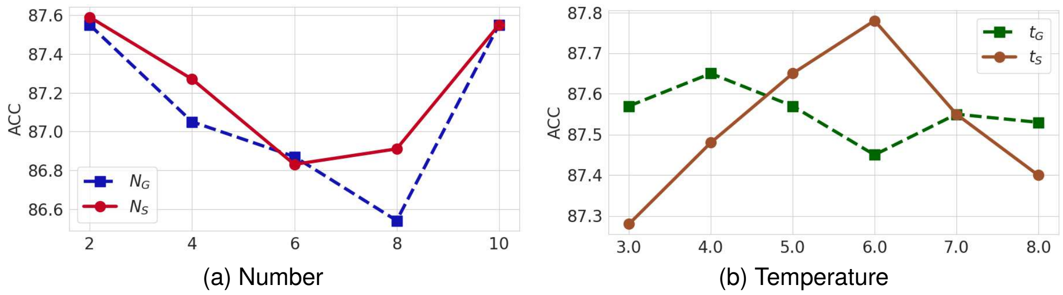


Figure 3: Hyperparameter Analysis

from the model, 'w/o LLM' refers to data augmentation using the PROTAUGMENT interpretation model, and 'w/o Retrieval' refers to the removal of the retrieval in-context prompt that leads to larger model cognition. We conduct ablation experiments on Liu57 and Reuters datasets.

The experimental results show that removing either mechanism will result in the degradation of model performance. If all improvements are removed (w/o $\mathcal{L}_G \& \mathcal{L}_S$), performance on 1-shot on Liu57 will drop by 6.97%, 3.93% reduction on Reuters 5-shot. Removing self-augmented unsupervised contrastive learning loss will result in a 6.25% performance reduction on Liu57 1-shot and if the proposed LLM-augmented algorithm is not used, the performance of the model is reduced by 1.54%. In addition, removing the retrieval-based in-context prompt method results in a 2.03% performance degradation on Reuters 1-shot.

3.6. Hyperparameter Analysis

In order to analyze some important parameters in unsupervised contrastive learning loss, we conducted hyperparameter analysis experiments. As shown in Figure 3, the lines in Figure 3 (a) respectively refer to the number of groups N_G in group-level unsupervised contrastive loss and the number of samples N_S in sample-level contrastive loss. The lines in Figure 3 (b) refer to the temperature t_G in group-level unsupervised contrastive learning loss and the temperature t_S in sample-level contrastive loss respectively. A lower temperature signifies a focus on challenging samples in contrastive learning, while a higher temperature emphasizes overall performance.

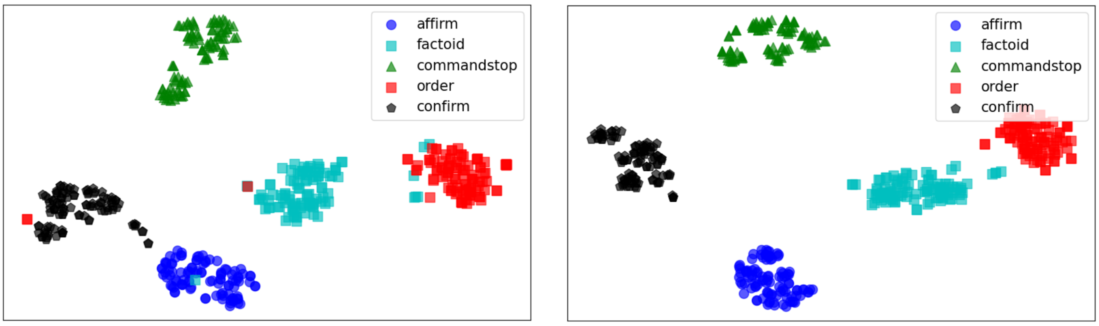
In subgraph (a), as the number of groups and samples increases, the model performance shows a trend of first declining and then improving, because the generated enhanced data introduces some noise and additional external information, which means that the model must either introduce as much information as possible more information, or introduce as little noise as possible. In subfigure (b), the influence of the temperature in the self-augmented unsupervised contrastive learning and the temperature in the external-augmented unsupervised contrastive learning on the model shows an opposite trend. This means that the self-augmented UCL is more focusing on some difficult-to-distinguish samples, external-augmented UCL pays attention to the overall overfitting problem.

3.7. Visual Analysis

Discriminating experiments of similar classes We conducted discriminative experiments on the Liu57 dataset to distinguish between similar classes. Specifically, for the 5-way 1-shot task's test set, we selected five similar classes ('affirm', 'factoid', 'commandstop', 'order', and 'confirm') and collected 100 samples per class. We employed the trained ContrasNet and LA-UCL model to obtain sample representations. We clustered the representations of ContrasNet and LA-UCL using the t-SNE algorithm (Van der Maaten and Hinton, 2008), and the resulting clusters are depicted in Figure 4.

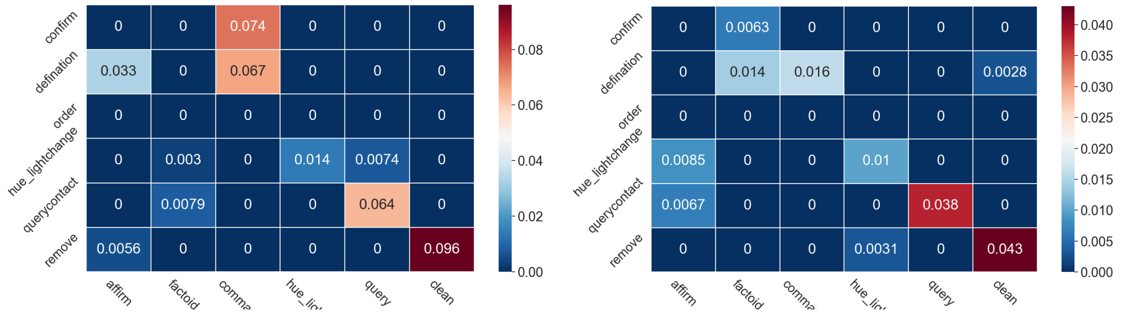
The experimental results demonstrate that our model is capable of effectively discriminating samples from similar classes, while some sample confusion may occur in ContrasNet. This validates our improvement on group-level unsupervised contrastive learning based on ContrasNet. By observing Figure 4 (a), it is evident that there are varying degrees of confusion between the 'factoid' class and the 'order' class, the 'confirm' class and the 'order' class, as well as the 'affirm' class and the 'factoid' class. This demonstrates that internal retrieval guides the LLM's cognition and generates more discriminative data-augmented samples. Furthermore, at the group-level, we incorporate an unsupervised contrastive learning mechanism between the current batch group and the base class groups in the training set, thereby effectively addressing the challenge of identifying similar classes within the contrastive learning framework.

Error Analysis To evaluate whether our algorithm enhances the classification performance of the model on similar classes, we performed an error rate analysis experiment. Specifically, we chose



(a) ContrasNet (b) LA-UCL

Figure 4: Discriminative experiment on Liu57.



(a) ContrasNet (b) LA-UCL

Figure 5: The error analysis experiment.

two sets of categories: ['affirm', 'factoid', 'commandstop', 'hue_lightdim', 'query', 'clean'] and ['confirm', 'defination', 'order', 'hue_lightchange', 'querycontact', 'remove']. The two sets contain closely related categories in corresponding positions, such as 'remove' and 'clean', 'query', and 'querycontact'. Additionally, there are certain similarities between categories in different corresponding positions (e.g., 'affirm' and 'definition'). In the form of a heat map, we show the probability that samples of one class is incorrectly predicted to be of another class.

As shown in Figure 5, the heat map demonstrates that our error rate for similar classes is significantly lower compared to the ContrasNet model. For instance, the error rate between 'clean', 'remove' in LA-UCL is 4.3%, whereas an error rate of ContrasNet is 9.6%. The experiment confirms that LA-UCL effectively enhances the model's ability to classify few samples by the self-augmented unsupervised contrastive learning.

4. Related Work

Generative Augmented Method. Data augmentation technology is crucial to few-shot solutions such as meta-learning and contrastive learning, and the challenge of generative data augmentation is how to generate high-quality and reliable data. Based on the EDA (Wei and Zou, 2019) and back-translation (Sennrich et al., 2016), PROTAUGMENT proposed a short text paraphrasing model, which can generate multiple paraphrases of the original text as data augmentation (Dopierre et al., 2021). Some studies use a prompting-based approach to generate labeled data from off-the-shelf language models (LMs) to optimize model performance in few-shot tasks (Dai et al., 2023; Sahu et al., 2022; Hou et al., 2022).

Contrastive Learning In text classification few-shot technology, contrastive learning can bring text representations belonging to the same category closer and push away text representations belonging to different categories, which has more advantages than work based on meta-learning (Chen et al., 2022). (Grover et al., 2022) propose a task-aware contrastive learning framework and (Sun, 2023) propose a novel contrastive consistency to improve model performance and refine sentence representation. UEFTC (He et al., 2023) proposes contrastive learning from Uncertainty relations to address uncertainty estimation for few-shot text classification. However, in solving the over-fitting problem, the performance of contrastive learning is limited by data augmentation technology; in addition, the lack of interaction with the base classes results in some difficult-to-discriminate samples not being fully learned.

5. Conclusion

In this paper, we propose the LA-UCL model, which is based on the Large Language Model (LLM) within the contrastive learning framework to achieve efficient data augmentation. Specifically, the self-augmented contrastive learning module guides the LLM in generating more discriminative augmented data by retrieving similar but different categories. Moreover, the external-augmented contrastive learning modules utilize web knowledge retrieval to expand the sample space and enable the LLM to generate more diverse data. Both modules introduce corresponding contrastive loss functions, which improves the model's ability to distinguish difficult samples and alleviates the overfitting problem. In the experiments, LA-UCL achieves optimal experimental performance on six datasets.

In the future, we will consider improvements in the following three areas. The first is sequence expansion based on LLM. For LLM with limited modeling length, it is difficult to effectively augment data for some tasks with extremely long sequences. Therefore, it is necessary to consider how to perform sequence expansion without training or with a small amount of training. The second is to explore the impact of data segmentation on few-shot learning. Third, our retrieval-based in-context prompt guides the cognition of LLM, but we will consider further training based on retrieval technology and develop soft prompting technology more suitable for data augmentation tasks.

6. Acknowledgements

This work is supported in part by the Natural Science Foundation of China (grant No.62276188), TJU-Wenge joint laboratory funding.

7. Reference

Yujia Bao, Menghua Wu, Shiyu Chang, and Regina Barzilay. 2020. Few-shot text classification with distributional signatures. In *International Conference on Learning Representations*.

Kaidi Cao, Maria Brbic, and Jure Leskovec. 2021. Concept learners for few-shot learning. In *International Conference on Learning Representation (ICLR)*.

Iñigo Casanueva, Tadas Temčinas, Daniela Gerz, Matthew Henderson, and Ivan Vulić. 2020. Efficient intent detection with dual sentence encoders. In *Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI*, pages 38–45, Online. Association for Computational Linguistics.

Junfan Chen, Richong Zhang, Yongyi Mao, and Jie Xu. 2022. Contrastnet: A contrastive learning framework for few-shot text classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 10492–10500.

Haixing Dai, Zhengliang Liu, Wenxiong Liao, Xiaoke Huang, Zihao Wu, Lin Zhao, Wei Liu, Ninghao Liu, Sheng Li, Dajiang Zhu, et al. 2023. Chataug: Leveraging chatgpt for text data augmentation. *arXiv preprint arXiv:2302.13007*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Thomas Dopierre, Christophe Gravier, and Wilfried Logerais. 2021. Protaugment: Unsupervised diverse short-texts paraphrasing for intent detection meta-learning. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 2454–2466.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning*, pages 1126–1135. PMLR.

Tianyu Gao, Xu Han, Zhiyuan Liu, and Maosong Sun. 2019. Hybrid attention-based prototypical networks for noisy few-shot relation classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 6407–6414.

Ruiying Geng, Binhua Li, Yongbin Li, Xiaodan Zhu, Ping Jian, and Jian Sun. 2019. Induction networks for few-shot text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3904–3913.

Ramansh Grover, Rachit Saksena, and Rahul Katarya. 2022. Improving few-shot text classification with task-aware contrastive learning. In *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)*, pages 1–8.

Hongyu Guo, Yongyi Mao, and Richong Zhang. 2019. Augmenting data with mixup for sentence classification: An empirical study. *arXiv preprint arXiv:1905.08941*.

Chengcheng Han, Zeqiu Fan, Dongxiang Zhang, Minghui Qiu, Ming Gao, and Aoying Zhou. 2021. Meta-learning adversarial domain adaptation network for few-shot text classification. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1664–1673.

Jianfeng He, Xuchao Zhang, Shuo Lei, Abdulaziz Alhamadani, Fanglan Chen, Bei Xiao, and Chang-Tien Lu. 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*, pages 698–710.

Yutai Hou, Hongyuan Dong, Xinghao Wang, Bohan Li, and Wanxiang Che. 2022. Metaprompting: Learning to learn better prompts. *arXiv preprint arXiv:2209.11486*.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. 2020. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673.

Diederik P Kingma and Jimmy Lei Ba. Adam: Amethod for stochastic optimization.

Brenden M Lake, Tal Linzen, and Marco Baroni. 2019. Human few-shot learning of compositional instructions. In *41st Annual Meeting of the Cognitive Science Society: Creativity+ Cognition+ Computation, CogSci 2019*, pages 611–617. The Cognitive Science Society.

Xingkun Liu, Arash Eshghi, Pawel Swietojanski, and Verena 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*, pages 165–183. Springer.

Shikhar Murty, Tatsunori B Hashimoto, and Christopher D Manning. 2021. Dreca: A general task augmentation strategy for few-shot natural language inference. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1113–1125.

Gaurav Sahu, Pau Rodriguez, Issam Laradji, Parmida Atighehchian, David Vazquez, and Dzmitry Bahdanau. 2022. Data augmentation for intent classification with off-the-shelf large language models. In *Proceedings of the 4th Workshop on NLP for Conversational AI*, pages 47–57, Dublin, Ireland. Association for Computational Linguistics.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96, Berlin, Germany. Association for Computational Linguistics.

Jake Snell, Kevin Swersky, and Richard Zemel. 2017. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30.

Liwen Sun. 2023. Few-shot text classification with dual contrastive consistency training.

Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of machine learning research*, 9(11).

Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. 2020. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, 53(3):1–34.

Jason Wei and Kai Zou. 2019. EDA: Easy data augmentation techniques for boosting performance on text classification tasks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6382–6388, Hong Kong, China. Association for Computational Linguistics.

Menghua Wu et al. 2020. *Few-shot text classification with distributional signatures*. Ph.D. thesis, Massachusetts Institute of Technology.

Mo Yu, Xiaoxiao Guo, Jinfeng Yi, Shiyu Chang, Saloni Potdar, Yu Cheng, Gerald Tesauro, Haoyu Wang, and Bowen Zhou. 2018. Diverse few-shot text classification with multiple metrics. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1206–1215.