Supplementary: Semi-Supervised Few-shot Learning via Multi-Factor Clustering

Jie Ling[1]*, Lei Liao[1]*, Meng Yang[1,2]†, Jia Shuai[1]
[1] School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou, China
[2]Key Laboratory of Machine Intelligence and Advanced Computing (SYSU), Ministry of Education
{lingj8, liaolei3, shuaij}@mail2.sysu.edu.cn,yangm6@mail.sysu.edu.cn

In this supplementary document, we discuss the further results and analysis of our cluster-FSL:

1. The impact of different number of extra unlabeled samples. (section 1)

2. The impact of ϵ in our Multi-Factor Clustering. (section 2)

3. The quality of pseudo-labels obtained by different methods. (section 3)

4. The efectiveness and uniqueness of our cluster-FSL. (section 4)

5. The ablation experiment for embedding propagation trick. (section 5)

1. Different Extra Unlabeled Samples

On the miniImageNet data set, we compared the effects of different numbers of unlabeled samples on the experimental results. And the results are shown in Table 1 and Table 2. Specifically, when the number of unlabeled samples is zero, our cluster-FSL does not use extra unlabeled samples. Among them, we regard the data in the support set as labeled data and perform model fine-tuning on the data in the query set. And we compare the cluster-FSL using only Multi-Factor Clustering (our-MFC) and using only Label Propagation (our-LP). From the experimental results, when only the MFC model is used to obtain the soft label and the model is fine-tuned, the accuracy of the model is higher than that when only the label propagation is used. Besides, it can be seen from the tables that when the number of introduced unlabeled samples increases, the accuracy of our model in each task on the miniImageNet dataset increases, which indicates the number of unlabeled data is vital to our method. Without extra unlabeled data, our method only obtain 71.93% accuracy in the 5-way 1-shot, but we can achieve 83.47% accuracy with 200 unlabeled data, which improves the accuracy by 11.54%.

*Equal contribution, † Corresponding author.

Table 1. The average classification accuracy of 1000 few-shot tasks in the 5-way 1-shot and 5-way 5-shot scenarios of the cluster-FSL model with different number of extra unlabeled samples. The backbone is WRN-28-10 and the dataset is miniImageNet. "Num_U" represents the number of unlabeled samples.

Methods	Num_U	5-way 1-shot	5-way 5-shot
our-LP	0	70.15%	83.03%
our-MFC	0	71.93%	83.29%
our	20	77.86%	86.84%
our	50	79.93%	88.24%
our	100	82.63%	89.16%
our	200	**83.47%**	**89.89%**

Table 2. The average classification accuracy of 1000 few-shot tasks in the 5-way 1-shot and 5-way 5-shot scenarios of the cluster-FSL model with different number of extra unlabeled samples. The backbone is ResNet-12 and the dataset is miniImageNet. "Num_U" represents the number of unlabeled samples.

Methods	Num_U	5-way 1-shot	5-way 5-shot
our-LP	0	64.93%	78.94%
our-MFC	0	67.29%	78.95%
our	20	72.58%	82.89%
our	50	74.83%	84.39%
our	100	77.81%	85.55%
our	200	**78.31%**	**86.25%**

2. Impact of Parameter ϵ

The parameter ϵ is used in Eq.(2), and is a constant to regularize the representation. It can be seen from Table 3 that the effect of ϵ is subtle. When the value of ϵ is set to 0.01, our method can achieve the best performance.

3. Quality of pseudo-labels

We use the accuracy of the pseudo-labels as a measure of the quality of the pseudo-labels. We conduct the experiments on miniImageNet dataset and tieredImageNet dataset

Table 3. The average classification accuracy of 1000 few-shot tasks in the 5-way 1-shot and 5-way 5-shot scenarios of the cluster-FSL model with different ϵ, which the backbone is WRN-28-10 and the dataset is miniImageNet.

ϵ	0.001	0.01	0.1	1
ACC	82.59%	**82.63%**	82.62%	82.58%

Table 4. Comparison of the accuracy(%) of the generated pseudo-labels, where these methods use WRN-28-10 as the backbone.

Methods	miniImageNet		tieredImageNet	
	1-shot	5-shot	1-shot	5-shot
Label Propagation	70.30	84.14	77.94	87.82
K-Means	81.39	87.62	85.00	88.36
Soft K-Means	81.50	87.69	85.00	88.56
MFC	**81.90**	**88.38**	**85.44**	**89.90**

Table 5. Comparison of the accuracy(%) of different cluster methods, where these methods use WRN-28-10 as the backbone.

Methods	miniImageNet	
	1-shot	5-shot
Label Propagation	79.81±0.67	87.99± 0.37
K-Means	82.14± 0.80	88.05± 0.40
Soft K-Means	82.18± 0.79	88.10± 0.40
MFC	**82.63± 0.79**	**89.16± 0.35**

in the 5-way 1-shot and 5-way 5-shot scenes. And the WRN-28-10 is used as the backbone of model in the experiments. In Table 4, the accuracy of pseudo-labels is the average accuracy of 1000 tasks. we observe that the MFC has the highest accuracy of the pseudo-labels under all settings. Moreover, the MFC can obtain more accurate pseudo-labels in the 5-shot scene. Specifically, the accuracy of pseudo-labels is 89.90% in 5-shot scene on the tieredImageNet dataset, which is 2.08%, 1.54% and 1.34% higher than Label propagation, K-means and Soft K-Means, resperctively. The results show that our proposed MFC can indeed obtain pseudo-labels with higher accuracy because MFC uses the reconstruction error of instead of Euclidean distance as the metric. Compared with the 1-shot, more labeled samples in the 5-shot can make the dictionary more complete and the reconstruction error will be more accurate.

4. The Effectiveness and Uniqueness of our Cluster-FSL

we compare different clustering methods alone in cluster-FSL to prove the effectiveness and uniqueness of our method, as shown in Table 5. We conduct the experiments on miniImageNet dataset in the 5-way 1-shot and 5-way 5-shot scenes. And the WRN-28-10 is used as the backbone of model in the experiments. Using MFC to predict the unlabeled data and query data at the same time, we got the 82.63±0.79% accuracy rate in the miniImageNet 1-shot setting with WRN-28-10. The results show that the performance of the model with MFC alone is higher than that using Label Propagation (LP) alone. When MFC and LP are combined, there is a higher accuracy rate. The reason is that MFC should use labeled samples to construct the dictionary. However, the pseudo labels of expanded support-set data have noise, which is detrimental to MFC. Thus using LP to infer the labels of query samples in the testing phase is reasonable.

Table 6. Ablation experimental results(%) for the EP trick, where these methods use WRN-28-10 as the backbone.

Datasets	Methods	1-shot	5-shot
mini	w EP	82.63±0.79	89.16±0.35
	w/o EP	**83.68±0.75**	**89.27±0.34**
	w/o EP inductive	83.66±0.75	89.25±0.34
tiered	w EP	85.74±0.76	90.18±0.43
	w/o EP	**86.64±0.73**	**90.61±0.42**
	w/o EP inductive	86.62±0.73	90.59±0.42

5. The Ablation Experiment for Embedding Propagation

We conducted experiments without using embedding propagation (EP) in the test phase and our method has further improvement as shown in the Table 6. The reason for the improvement may be that MFC uses the reconstruction error as the metric and EP uses the Euclidean distance, which results in performance degradation. Although our cluster-FSL is in the transductive setting, we try to use the inductive setting for testing, that is, only one query sample is inputted at a time, and then its category is predicted, without using the information of all query samples. The results of Table 6 show that our cluster-FSL is also robust in such a setting. This is because the unlabeled samples already provide enough distribution information, and more query samples information can be the icing on the cake.