Kernel Relative-prototype Spectral Filtering for Few-shot Learning

Tao Zhang[1] and Wu Huang[2]
1 Chengdu Techman Software Co., Ltd. 2 Sichuan University

Introduction

Few-shot learning performs classification tasks and regression tasks on scarce samples. As one of the most representative few-shot learning models, Prototypical Network represents each class as sample average, or a prototype, and measures the similarity of samples and prototypes by Euclidean distance. In this paper, we propose a method called Deep Spectral Filtering Networks (DSFN) aiming to better estimate relative prototypes, or the difference between prototypes and query samples (Fig. 1). Euclidean distance can not fully capture the information of intra-class difference as it is applied to measure the difference between sample and prototype. Thus, the main components of the inner class distance are taken into ac count in the similarity assessment between the sample and the prototype, which to some extent interferes with this assessment. In our approach, the influence of these components is weakened via spectral filtering.

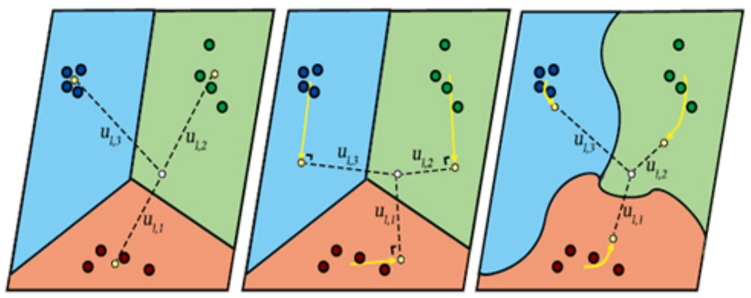

Fig.1 Comparison of the relative-prototypes (dotted line) in the Prototypical Networks (ProtoNet), Deep Subspace Networks (DSN) and the proposed DSFN. Left: ProtoNet in Euclidean space; Middle: DSN in Euclidean space; Right: DSFN in a reproducing kernel Hilbert space. For them, $u_{l,1}$,$u_{l,2}$ and $u_{l,3}$ are the relative-prototypes with class 1(brown), class 2(green) and class 3(blue), respectively.

Methods

The similarity of sample-pairs can be measured using the square of distance between the relative-prototype and original point in RKHS

$$d^2_{l,c}(\lambda, \boldsymbol{S}_c, \boldsymbol{q}_l) = (\boldsymbol{\alpha}_{l,c}(\lambda))^T \tilde{\boldsymbol{K}}^c_{ss} \boldsymbol{\alpha}_{l,c}(\lambda) + \boldsymbol{I}_n^T \tilde{\boldsymbol{K}}^{l,c}_{qq} \boldsymbol{I}_n - 2(\boldsymbol{\alpha}_{l,c}(\lambda))^T \tilde{\boldsymbol{K}}^{l,c}_{qs} \boldsymbol{I}_n$$

The probability of the sample q_l in query set belonging to class c is

$$P_{\omega,\theta}(Y = c|\boldsymbol{q}_l) = \frac{\exp\left(-\zeta d^2_{l,c}(\lambda, \boldsymbol{S}_c, \boldsymbol{q}_l)\right)}{\sum_{c=1}^{C} \exp\left(-\zeta d^2_{l,c}(\lambda, \boldsymbol{S}_c, \boldsymbol{q}_l)\right)}$$

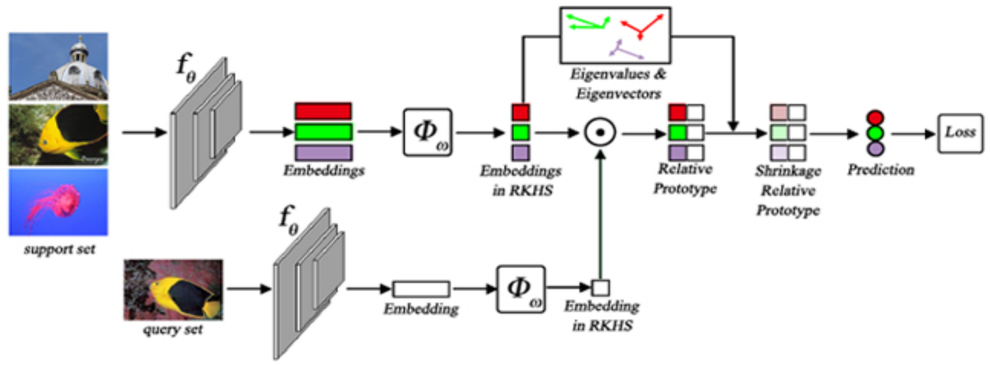

Fig. 2 The overview of our proposed approach. The features of support set and query set extracted from f_θ are mapped into RKHS by the function φ_ω. The relative prototypes are shrunk based on the eigenvalues and eigenvectors from the support set.

Results

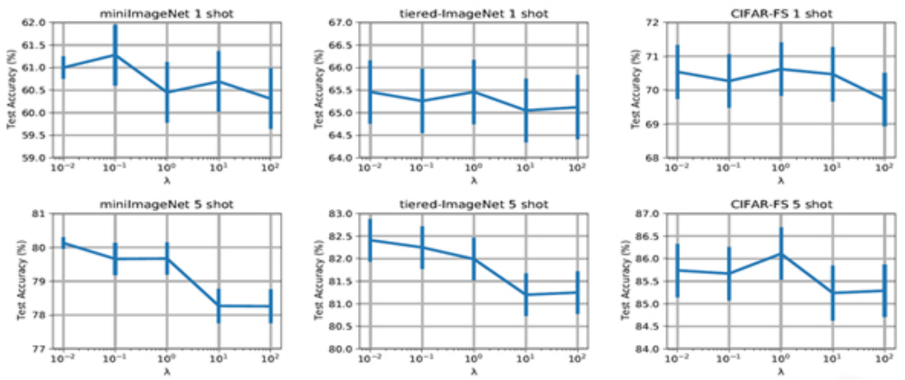

Fig.3 Test accuracies on few-shot classification tasks from the proposed DSFN against different values of shrinkage parameter, where ResNet-12 is used.

Results

Table 1: Accuracy comparison (%) between models with and without shrinkge on different datasets for 5-way 1-shot and 5-way 5-shot classification tasks, w S: with shrinkge, w/o S: without shrinkge.

Dataset	Kernel	w S	w/o S	1-shot	5-shot
*mini*ImageNet	identity		✓	60.14 ± 0.67	77.65 ± 0.52
	identity	✓		61.00 ± 0.25	80.13 ± 0.17
	RBF		✓	58.74 ± 0.65	79.18 ± 0.46
	RBF	✓		59.43 ± 0.66	79.60 ± 0.46
tiered-ImageNet	identity		✓	65.05 ± 0.72	81.14 ± 0.55
	identity	✓		65.46 ± 0.70	82.41 ± 0.53
	RBF		✓	64.23 ± 0.70	82.07 ± 0.53
	RBF	✓		64.27 ± 0.70	82.26 ± 0.52
CIFAR-FS	identity		✓	70.54 ± 0.82	85.30 ± 0.59
	identity	✓		70.62 ± 0.79	86.11 ± 0.58
	RBF		✓	71.18 ± 0.73	86.09 ± 0.47
	RBF	✓		71.28 ± 0.70	86.30 ± 0.46

Conclusion

In this work, we propose a framework called DSFN for few-shot learning. In this framework, one can represent the similarity between a query and a prototype as the distance after spectral filtering of support set in each class in RKHS. DSFN is an extension of some mainstream methods, e. g., ProtoNet and DSN, and with appropriate filter function, the framework of DSFN can be embodies as those methods. In addition, we also showed that in this framework, one can explore new methods by applying diverse forms, e. g., Tikhonov regularization, as the filter function, and diverse forms of kernels. Several experiments verified the effectiveness of various specific forms of the proposed DSFN. Future works should take a closer look at the selection of filter function and the role of shrinkage parameter in the proposed framework.