Research on the Few-Shot Learning Based on Metrics

Yican Shen[1], [*]

[1]Faculty of Electronic Information and Electric Engineering Dalian University of Technology, Dalian City, Liaoning Province, P.R.C., 116024

ABSTRACT: Deep learning has been rapidly developed and obtained great achievements with a data-intensive condition. However, sufficient datasets are not always available in practical application. In the absence of data, humans can still perform well in studying and recognizing new items while it becomes a hard task for the computer to learn and generate from a small dataset. Thus, researchers are increasingly interested in few-shot learning. The purpose of few-shot learning is to allow computers to carry out unknown tasks with a few examples. Recently, effective few-shot models have frequently been designed using transfer learning approaches, with the metric method being an important branch in transfer learning. This article reviews the metric methodologies for few-short learning, analyzing the development of the metric based few-shot learning in the following three categories: traditional metric methods, relation network based metric methods and graph based metric methods. Then it compares the effectiveness of those models on a representative dataset and illustrates the feature of each category. Finally, it discusses the potential future research fields.

1. INTRODUCTION

The main application area of few-shot learning techniques is currently image recognition. The metric method is an important branch of transfer learning-based few-shot image recognition. Transfer learning is the use of old knowledge to learn new knowledge, with the main goal of transferring what has been learned to a new domain [1]. Transfer learning requires that the source and target domains are related in some way, so that the knowledge and features learned in the source can help train a classification model in the target domain, thus enabling the transfer of knowledge between different domains. In transfer learning, the dataset is split into 3 parts training set, support set and query set. The training set contains a large amount of labeled data; the support set is the training sample in the target domain, which contains a small amount of annotated data; the query set is the test sample in the target domain. For few-shot tasks, the models are generated from the source domain with a huge dataset and the target samples are embedded in the model to make the classification.

The work of the metric learning program is a similarity measure procedure. The datasets are embedded into the feature vector space. Then the metric part compares the similarity of the input data to the training dataset and selects the category with the highest similarity 2. This progress transfers the knowledge of the dataset to the unknown categories.

In image recognition tasks, the few-shot learning task is also called as the C-way K-shot tasks. The models have generally tested their effectiveness on the 5-way 1-shot and 5-way 5-shot tasks on the datasets. Frequently used datasets are Omniglot, including 1623 handwriting images of 50 unknown letters; miniImageNet, including 60000 images of 100 categories; tiered Image Net, a larger dataset with 608 categories.

2. METRIC MODELS DEVELOPMENT

The metric and similarity based one-shot classification model was first introduced in 2015 by using the Siamese neural network 3. Then Snell et al proposed a prototypical network in 2017 to solve the few-shot tasks 4. Those two models both traditional metric functions like Euclidean distance in the embedding space to measure the similarity. Research after the relation network starts adjusting the traditional metric part into neural networks 5 9.

2.1. Traditional Metric Methods

The Siamese neural network embeds two input samples into the feature vector space using two identical CNN networks, then calculates the distance using the Euclidean distance to reflect the similarity of the two samples [3]. The training progress is to minimize the distance of samples in the same category and maximize the distance of different ones. Matching network raised in 2016 uses LSTM to map the labeled and unknown samples on the embedding space and assesses similarity on kernel density estimation 6. Related research improves the embedding methods and similarity parts based on the Siamese neural network principle.

Those metric model models all work for one-shot learning and recognition tasks and cannot improve much

---

*Corresponding author. Email: syc1061@mail.dlut.edu.cn

performance when the sample increases to five. Prototypical network improves the performance of the model on few-shot tasks by introducing the prototype of each category. The network looks for an embedding space where samples (more than two) are close to their prototype and to each other 4. New samples are recognized by the Euclidean distance between them and the prototypes. The introduction of the prototype approves the metric models as a targeted solution on few-shot problems. Ren et al refine the prototypical network with semi-supervised methods 7. A basic prototypical network calculates the prototype with the only labeled dataset, which obtains an imprecise result. This work adds unlabeled data into the training dataset to advance the border of the categories. This work uses soft K-means to improve the prototypes' accuracy and variant ways to eliminate the outline points. Gao et al. improve the classification accuracy on conditions with noisy datasets by combining the attention-based prototypical network with the attention mechanism [8]. All the models above use the traditional metric function to compare the similarity. These metric functions are easily understood and applied while the accuracy performance of the recognition is not satisfying in some missions.

2.2. Relation Network-Based Metric Methods

The relation network first replaces the traditional metric function with the neural networks [5]. The relation part of this model is a ReLU unit. The relation network model can deal with both few-shot tasks and zero-shot tasks. In few-shot tasks, feature vectors of the input sample and vectors of the possible categories are processed in the ReLU unit to calculate the similarity. Then, the input sample is classified by the similarity with all the categories like traditional metric models do; in zero-shot tasks, the semantic feature vector of each category and a new embedding function are used, to obtain the feature mapping for this new category, and the rest of the procedure is the same as above. Based on a relation network, a deep comparison network divides the embedding part and the relation part into series 16. Each embedding unit is paired with a relation unit. The relational module uses the representation of the corresponding embedding module to compute a nonlinear measure to score the matches. To ensure that all features of the embedding unit are used, the relation units are monitored in depth.

The simple relation network methods focus on the first-order statistics. Covariance metric network utilizes the covariance representation and covariance matrix,

where the covariance representation is used to capture second-order statistical information and the covariance matrix is used to measure the consistency of the distribution between the query sample and the new category [9] [10]. He et al use Memory-Augmented Relation Network which stores sample information to enhance the samples and uses undirected graphs instead of CNN to train the embedding function for support set samples information propagation to improve the embedding ability to obtain better feature representation [11]. Kang et al propose to learn relational patterns from self-correlation within an image representation and cross-correlation between two image representations to form relational embedding [14].

2.3. Graph Based Metric Methods

Garcia et al use graph neural network (GNN) to classify images in few-shot tasks [12]. The key change in the GNN approach is the metric module, using a graph network to train the distance metric function, while it uses a graph network to train the distance metric function by traversing the undirected graph to accomplish sample metrics and classification. In the GNN, each sample is considered as a node in the graph. The model learns not only the embedding vector of each node, but also the embedding vector of each edge. Convolutional neural networks embed all samples into the vector space, connect the sample vectors with the label vectors and input them into the graph neural network to build the connected edges between each node; then the node vectors are updated by the graph convolution. The node vector is updated by the node vector, and then the edge vector is updated by the node vector, which constitutes a deep graph neural network.

Kim et al considered another aspect to classify the edges in the graph 13. Kim et al took another perspective and proposed Edge labeling graph neural network (ELGNN). In contrast to GNN, ELGNN classifies the edges of the graph, mainly by using the intra-class similarity and inter-class difference of nodes in the edges for iterative to drive the update of nodes, and explicitly construct the inter-sample metric to complete the sample classification 15.

3. EXPERIMENTS AND RESULTS

This article chooses the mini ImageNet to test the performance on image classification of the typical model above. All the results are chosen on 5-way 1-shot and 5-way 5-shot tests.

Table 1 Comparison of classify precision in mini ImageNet

Approaches	1-shot	5-shot
Matching Nets 2	43.56 ± 0.84%	55.31 ± 0.73%
Prototypical Nets 8	49.42 ± 0.78%	68.20 ± 0.66%
Semi-supervised Prototypical Nets 7	50.41 ± 0.31%	64.39 ± 0.24%
Relation Net 5	50.44 ± 0.82%	65.32 ± 0.70%
MRN 11	57.83 ± 0.69%	71.13 ± 0.50%
CovaMNet 9	51.19±0.76%	67.65 ± 0.63%
GNN [12]	50.33±0.36%	66.41±0.63%
EGNN [13]	-	76.37%

From the results, we can find out that all the methods perform better on 5-shot tasks than 1-shot tasks while models performances improve differently with the increment of the sample. As a few-shot focused model, Prototypical Nets has a large increment of precision on 5-shot tasks. In some cases, such as noisy datasets, neural network-based metric models such as deep comparison network and MRN can be advantageous.

4. DISCUSSION

This article describes the development process and features of different metric-based few-shot learning methods. It analyzes metric-based learning models into three categories including traditional metric methods, relation network based methods and graph based methods.

However, it seems that the classification of models on few-shot learning is not certain. In some works, the metric methods and graph neural networks are both branches of transfer learning. And some people categorize many of the methods that follow traditional metric models as meta-learning methods. The definition and classification of meta-learning, metric learning, and embedding methods are still controversial. This article is an introduction and analysis based on one of the more commonly accepted classifications.

Despite the different definitions, various models are dedicated to solving the few-shot image recognition problem using different approaches. Traditional methods first introduce a new concept in transfer learning methods [3][6]. Relation networks and graph neural networks develop based on traditional methods and each perform better in specific areas [5] [13].

Metric-based methods for few-shot learning are developing rapidly. Recognition accuracy and model generalization are improving. Transfer learning and metric-based few-shot learning still deserves more in-depth study. In the future, simply optimizing traditional metric learning models may not yield good results. Focusing on developing the recent models or combining the relation networks, graph networks, and other meta-learning methods both likely have the potential to lead to further development of metric-based few-shot learning.

5. CONCLUSION

Due to the small sample size or annotated samples in some real-world domains, and the sample annotation work can be time-consuming and labor-intensive, in recent years, few-shot learning has become a key concern. This article

introduces the development of metric based few-shot learning in three categories, mainly focusing on the metric function in those models. This field develops for one-shot learning to few-shot learning to zero-shot learning. Traditional metric function has limited ability in comparing the similarity, and those models enhance the ability of classification only based on the embedding part. The indications of similarity of the feature vectors are more diverse with neural network and metric neural network-based metric functions, and there are various models found on relation networks for the elemental and specific few-shot classification tasks.

On the metric function part, it is a difficulty to obtain a large improvement in the classification accuracy on few-shot tasks by continuing to use the traditional distance function-based method. Thus, focusing on the neural network metric function design is a practical field of the few-shot learning. However, the poor interpretability of the neural network makes it hard to design a network to compare the similarity, which means many attempts must be made to improve.

Graph based few-shot metric models proposed later than others. However, graph neural networks are highly interpretable and have relatively good performance, which means graph neural network improvement a feasible direction on few-shot learning. On those few shot tasks, research into how to design the graph network structure, node update function, and edge update function may result in significant improvements in model performance. In general, few-shot tasks in the field of image classification has been studied in great depth. Classifying the categories of existing datasets can obtain a not bad accuracy, but there is still a big gap compared with human classification in terms of accuracy and universality. A better embedding model to map the feature of samples and a better metric function to compare the similarity is always the target of metric based few-shot learning model.

REFERENCE

1. Zhuang, Fuzhen, et al. "A comprehensive survey on transfer learning." Proceedings of the IEEE 109.1 (2020): 43-76.

2. Vinyals, Oriol, et al. "Matching networks for one shot learning." Advances in neural information processing systems 29 (2016): 3630-3638.

3. Koch, Gregory, Richard Zemel, and Ruslan Salakhutdinov. "Siamese neural networks for one-shot image recognition." ICML deep learning workshop. Vol. 2. 2015.

4. Snell, Jake, Kevin Swersky, and Richard S. Zemel. "Prototypical networks for few-shot learning." arXiv preprint arXiv:1703.05175 (2017).

5. Sung, Flood, et al. "Learning to compare: Relation network for few-shot learning." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

6. Vinyals, Oriol, et al. "Matching networks for one shot learning." Advances in neural information processing systems 29 (2016): 3630-3638.

7. Ren, Mengye, et al. "Meta-learning for semi-supervised few-shot classification." arXiv preprint arXiv:1803.00676 (2018).

8. Gao, Tianyu, et al. "Hybrid attention-based prototypical networks for noisy few-shot relation classification." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. No. 01. 2019.

9. Li, Wenbin, et al. "Distribution consistency based covariance metric networks for few-shot learning." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 33. No. 01. 2019.

10. Gao, Hang, et al. "Low-shot learning via covariance-preserving adversarial augmentation networks." arXiv preprint arXiv:1810.11730 (2018).

11. He, Jun, et al. "Memory-augmented relation network for few-shot learning." Proceedings of the 28th ACM International Conference on Multimedia. 2020.

12. Garcia, Victor, and Joan Bruna. "Few-shot learning with graph neural networks." arXiv preprint arXiv:1711.04043 (2017).

13. Kim, Jongmin, et al. "Edge-labeling graph neural network for few-shot learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

14. Kang, Dahyun, et al. "Relational Embedding for Few-Shot Classification." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

15. Kim, Jongmin, et al. "Edge-labeling graph neural network for few-shot learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

16. Zhang, Xueting, et al. "RelationNet2: Deep comparison columns for few-shot learning." arXiv preprint arXiv:1811.07100 (2018).

17. Zheng, Yan, et al. "Principal characteristic networks for few-shot learning." Journal of Visual Communication and Image Representation 59 (2019): 563-573.

18. Wang, Yaqing, et al. "Generalizing from a few examples: A survey on few-shot learning." ACM Computing Surveys (CSUR) 53.3 (2020): 1-34.