Multi-label Few-shot Learning with Semantic Inference (Student Abstract)

Zhen Wang, Yiqun Duan, Liu Liu, Dacheng Tao
The University of Sydney, Darlington, NSW 2008, Australia
zwan4121@uni.sydney.edu.au

Abstract

Few-shot learning can adapt the classification model to new labels with only a few labeled examples. Previous studies mainly focused on the scenario of a single category label per example but have not effectively solved the more challenging multi-label scenario, which has exponential-sized output space and low-data. In this paper, we propose a semantic-aware meta-learning model for multi-label few-shot learning. Our approach can learn and infer the semantic correlation between unseen labels and historical labels to quickly adapt multi-label tasks based on only a few examples. Specifically, features can be mapped into the semantic space via label embeddings to exploit the label correlation, thus structuring the overwhelming output space. We design a novel semantic inference mechanism for leveraging prior knowledge learned from historical labels, which will produce good generalization performance on new labels to alleviate the overfitting caused by low-data. Finally, empirical results show that the proposed method significantly outperforms the existing state-of-the-art methods on the multi-label few-shot learning tasks.

Introduction

Few-shot learning is proposed to facilitate deep learning systems to learn new concepts with very limited labeled data, which is experiencing rapid development and advancements. Most of the latest few-shot learning works (Finn, Abbeel, and et al. 2017; Antoniou, Edwards, and Storkey 2019) are based on meta-learning (*learning-to-learn*) paradigm. Meta-learning is a task-level learning framework that aims to accumulate knowledge from learning a large number of tasks and generalize the knowledge to learn new tasks effectively.

Nevertheless, most of the previous works only focus on the scenario where each example is associated with an exclusive single label, but ignore the more actual and challenging scenario, in which each example can be simultaneously associated with multiple labels. The key challenge of multi-label few-shot learning (ML-FSL) is the huge output space, where the number of possible label sets exponentially grows with the increasing number of category labels. The huge output space simultaneously denotes a much sparser learning target and will bring difficulties in model learning for ML-FSL. Moreover, considering the fact that few-shot learning requires the model to be optimized on a relatively small dataset (Antoniou, Edwards, and Storkey 2019), a severe over-fitting problem will arise, which would also be aggravated by the huge output space as well.

In this paper, we propose a gradient-based meta-learning framework, DEep Semantic InfeREnce Network (DESIRE-Net), to solve the ML-FSL problem. Different from previous meta-learning methods, we utilize word embedding vectors instead of one-hot vectors as our prediction output to map features into semantic space, which gives us a tighter learning object. Specifically, our approach trains a semantic-aware feature extractor that maps features into the semantic embedding space via label word vectors learned from unsupervised text corpora. That is, the semantic correlation across labels can be preserved by the base model, which helps to structure and regularize the overwhelming output space of ML-FSL. Furthermore, we propose a meta-learning framework with a semantic inference mechanism that can extract semantic features and exploit the correlation between novel labels and historical labels as prior knowledge to classify multiple labels only using a few examples effectively. The semantic inference mechanism has two functions: training better initialization parameters of the model by leveraging the knowledge learned from historical tasks; inferring the classification of novel labels according to the semantic correlation with historical labels. Experimental results suggest that with the help of semantic inference, our model achieves state-of-the-art performance on ML-FSL, and ablation studies validate each module's effectiveness.

Methodology

Problem Definition

ML-FSL aims to learn a model that can be well adapted to novel multi-label tasks using only a few annotated examples. To be specific, we are given a sufficient labeled training set associated with a base label set L_{base}. Meanwhile, we also have a testing set with a disjoint set of novel labels L_{novel}, where each label is associated with only a few labeled examples. The **goal** of ML-FSL is to obtain a good multi-label classifier for the novel labels only using a few labeled examples. The huge output space of multi-label prediction and very few labeled examples limit to train a supervised classification model. To this end, we propose a meta-learning framework, Deep Semantic Inference Network, to explore and exploit the semantic label correlation and meta-knowledge for model novel labels.

Deep Semantic Inference Network

Deep Semantic Inference Network (DESIRE-Net) consists of two main components: a semantic-aware feature extractor and a meta-learner with the semantic inference mechanism.

First, we train an semantic-aware feature extractor $\mathcal{F}$ with a output matrix W_{base} on the training set to embed features into semantic space, where W_{base} is assigned and fixed to base-label word vectors learned from unsupervised text corpora. Through training with cross-entropy loss on the training set, the features produced by feature extractor (i.e., $\boldsymbol{z}{=}\mathcal{F}(\boldsymbol{x})$) would be embedded into a semantic space that naturally makes the output space of ML-FSL tighter, thus alleviate the overwhelming output space problem.

Second, a semantic inference mechanism is designed for leveraging the prior knowledge learned from base labels to adapt new multi-label tasks. Since we have obtained features embedded into the semantic space, the output feature $\boldsymbol{z}$ would naturally be closer to those with similar semantic meanings. In that case, we propose to infer our predictions by proximity between $\boldsymbol{z}$ and novel label word vectors W_{novel}, and use the semantic correlation between base-label word vectors and novel-label word vectors. The semantic inference mechanism $\mathcal{I}$ takes the form:

$$\mathcal{I}(\boldsymbol{z}){=}\left(\gamma_1 \frac{\mathcal{W}_Z(\boldsymbol{z})}{\|\mathcal{W}_Z(\boldsymbol{z})\|_2} {+}\gamma_2 \frac{\boldsymbol{z}W_{base}}{\|\boldsymbol{z}W_{base}\|_2} W_I\right) W_{novel} \ , \quad (1)$$

where $\mathcal{W}_Z(\boldsymbol{z})$ denotes a nonlinear transformation for $\boldsymbol{z}$, and $\|\cdot\|_2$ denotes l_2 normalization which can eliminate the influence of the absolute magnitudes of semantic features and improve the robustness; W_I is a learnable matrix trained on different tasks to learn the deep correlation between novel and base labels; γ_1 and γ_2 are the learnable module factors.

We adopt a meta-learning training process. In each iteration, a few examples are sampled for training to compute and update the meta-learner parameters that will achieve the maximal possible performance on the new task.

Experiments

Datasets and Evaluation

We conduct experiments on a widely used multi-label dataset, Delicious[1], which has 983 labels and 16105 examples. We split the dataset into three parts where labels are mutually disjoint: training set including 600 labels, validation set including 175 labels, testing set including 200 labels. For an N-way K-shot testing setting, each testing task is sampled with N labels, and each label includes K labeled examples and 15 testing examples. The testing results were evaluated on AUC (Wang, Liu, and Tao 2020).

Result Analysis and Ablation Study

Table 1 illustrates the average performance of our model in comparison with meta-learning baselines: MAML (Finn, Abbeel, and et al. 2017), Reptile (Nichol, Achiam, and Schulman 2018), ATAML (Jiang et al. 2018), and MAML++ (Antoniou, Edwards, and Storkey 2019).

[1]http://mulan.sourceforge.net/

Method	5-way AUC		10-way AUC	
	1-shot	5-shot	1-shot	5-shot
pre-trained	67.5%	75.3%	73.0%	81.7%
MAML	75.9%	81.9%	79.0%	82.3%
Reptile	76.1%	81.6%	79.2%	84.2%
ATAML	83.1%	82.9%	83.0%	85.1%
MAML++	81.3%	84.0%	80.6%	84.4%
DESIRE-Net$\backslash l_2$	84.0%	88.5%	83.8%	86.1%
DESIRE-Net$\backslash\mathcal{F}$	80.1%	85.8%	78.6%	83.5%
DESIRE-Net$\backslash\mathcal{I}$	78.2%	84.3%	77.2%	83.8%
DESIRE-Net(ours)	**87.2%**	**90.5%**	**87.9%**	**88.7%**

Table 1: Comparing multi-label few-shot classification performance on Delicious. Ablation study: $\backslash l_2$ denotes without l_2 normalization; $\backslash\mathcal{F}$ denotes feature extractor is removed; $\backslash\mathcal{I}$ denotes semantic inference mechanism is removed.

The results consistently show that the proposed model DESIRE-Net outperforms the baselines on the AUC metric for both 1-shot and 5-shot, 5-way and 10-way. Our approach can extract semantic features and exploit the correlation between novel labels and base labels as prior knowledge, therefore, it can achieve better results in dealing with the problem of multi-label few-shot learning. Ablation studies as shown in Table 1 validate the effectiveness of each proposed module: l_2 normalization, semantic-aware feature extractor, and semantic inference mechanism.

Conclusion

We propose a deep semantic inference network for multi-label few-shot learning, which can quickly adapt multi-label classification tasks using only a few labeled examples. Experiments illustrate that our approach could achieve the best results amongst other meta-learning models by introducing our proposed semantic inference mechanism.

Acknowledgements

This work was supported in part by Australian Research Council Projects DP-180103424 and IH-180100002.

References

Antoniou, A.; Edwards, H.; and Storkey, A. 2019. How to train your MAML. In *ICLR*.

Finn, C.; Abbeel, P.; and et al. 2017. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *ICML*.

Jiang, X.; Havaei, M.; Chartrand, G.; and et al. 2018. On the Importance of Attention in Meta-Learning for Few-Shot Text Classification. *CoRR* abs/1806.00852.

Nichol, A.; Achiam, J.; and Schulman, J. 2018. On First-Order Meta-Learning Algorithms. *ArXiv* abs/1803.02999.

Wang, Z.; Liu, L.; and Tao, D. 2020. Deep Streaming Label Learning. In *ICML*.