Latent Task Adaptation with Large-scale Hierarchies

Yufeng Jiang

Abstract

Recent years have witnessed the success of large-scale image classification systems that are able to identify objects among thousands of possible labels. However, it is yet unclear how general classifiers such as ones trained on ImageNet can be optimally adapted to specific tasks, each of which only covers a semantically related subset of all the objects in the world. In this paper authors propose a novel probabilistic model that jointly identifies the underlying task and performs prediction with a linear-time probabilistic inference algorithm, given a set of query images from a latent task. Empirical results based on the ImageNet data showed significant performance increase over several baseline algorithms.

1. Introduction

Recent years have witnessed a growing interest in object classification tasks involving specific sets of object categories, such as fine-grained object classification [1, 2] and home object recognition in visual robotics. Existing methods in the literature generally describe algorithms that are trained and tested on exactly the same task. A dog breed classifier is trained and tested on dogs and a cat breed classifier done on cats, without the use of out-of-task images.

However, two observations may render this "one classifier per task" approach suboptimal. First, its known that using images of related tasks is often beneficial to build a better model for the general visual world [5], which serves as a better regularization for the specific task as well. Second, object categories in the real world are often organized in, or at least well modeled by, a nested taxonomical hierarchy (*e*.g. Figure 1), with classification tasks corresponding to intermediate subtrees in this hierarchy, and recent efforts on the ImageNet challenge [3, 4] have leveraged the use of large-scale data to learn such information. While it is reasonable to train separate classifiers for specific tasks.

2. A Generative Model for Task Adaptation

Formally, authors define a classification task to be a subset of all the possible object labels that are semantically related. During testing time, a number of query images are
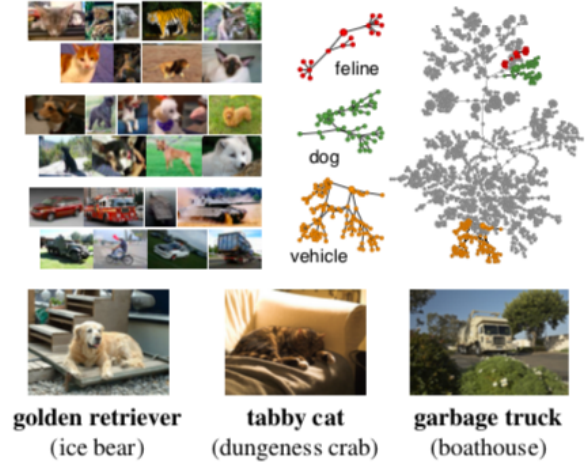

Figure 1. Top: Visualization of specific object classification tasks of interest in daily life, which are often subtrees in a large scale object taxonomy, *e*.g. the ImageNet hierarchy. Bottom: Adapting the ImageNet classifier allows us to perform accurate prediction (bold), while the original classifier prediction (in parentheses) suffers from a higher confusion.
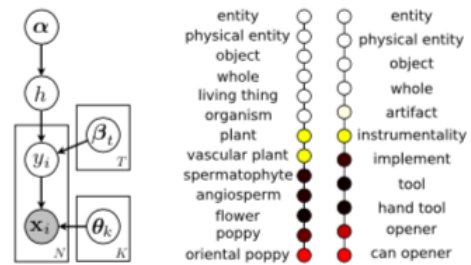

Figure 2. Left: the generative model for the latent task and corresponding query images. Right: the prior probabilities of the latent tasks from psychological study, along the path leading to the synsets oriental poppy and can opener respectively, with darker color indicating higher probability.

randomly sampled from the labels belonging to a task, and the learning algorithm needs to give predictions on these images. In this section they propose a probabilistic frame-

work that models the generation of latent tasks and the test time query images.

As stated in the previous section, they are interested in the scenario when the task is latent, *i.e.* only implicitly specified by the query images. We introduce two key components for modeling the generative process of query images: a latent task space that defines possible tasks and their probability, and a procedure to sample query images given a specific latent task. Specifically, we propose the graphical model in Figure 2 which generates a set of N query images when given T possible tasks and K object categories.

2.1. Latent Task Space

In this paper authors are mainly interested in modeling the frequency of various tasks in a general and large-scale setting1. Prior research on psychology and Bayesian generalization [6, 7] have shown that people favor basic-level concepts, which could be well modeled by a Erlang prior with respect to the size $|h|$ of each latent task, defined as the number of leaf nodes in the subtree rooted at the task:

$$P(h) = \alpha \propto (|h|/\sigma^2) exp(-|h|/\sigma) \qquad (1)$$

which favors medium-sized tasks corresponding to basic level concepts.

References

[1] R. Farrell, O. Oza, N. Zhang, V. I. Morariu, T. Darrell, and L. S. Davis. Birdlets: subordinate categorization using volumetric primitives and pose-normalized appearance. In *ICCV*, 2011. 1

[2] A. Khosla, N. Jayadevaprakash, B. Yao, and L. Fei-Fei. Novel dataset for fine-grained image categorization. In *CVPR FGVC workshop*, 2011. 1

[3] A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012. 1

[4] Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, L. Cao, and T. Huang. Large-scale image classification: fast feature extraction and svm training. In *CVPR*, 2011. 1

[5] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng. Self-taught learning: transfer learning from unlabeled data. In *ICML*, 2007. 1

[6] J. B. Tenenbaum and T. L. Griffiths. Generalization, similarity, and bayesian inference. *Behavioral and brain sciences*, 24(4):629–640, 2001. 2

[7] J. B. Tenenbaum, T. L. Griffiths, and C. Kemp. Theory-based bayesian models of inductive learning and reasoning. *Trends in cognitive sciences*, 10(7):309–318, 2006. 2