Pattern Recognition 67 (2017) 97–109

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Hierarchical learning of multi-task sparse metrics for large-scale image classification

Yu Zheng[a], Jianping Fan[b], Ji Zhang[c], Xinbo Gao[a,*]

[a] Department of Electronic Engineering, Xidian University, Xi'an 710071, PR China
[b] Department of Computer Science, University of North Carolina, Charlotte, NC 28223, USA
[c] Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an 710049, PR China

ARTICLE INFO

Article history:
Received 20 July 2016
Revised 18 December 2016
Accepted 24 January 2017
Available online 3 February 2017

Keywords:
Hierarchical multi-task sparse metric learning
Visual tree
Large-scale image classification

ABSTRACT

In this paper, a novel approach is developed to learn a tree of multi-task sparse metrics hierarchically over a visual tree to achieve a fast solution to large-scale image classification, where an enhanced visual tree is first learned to organize large numbers of image categories hierarchically in a coarse-to-fine fashion. Over the visual tree, a tree of multi-task sparse metrics is learned hierarchically by: (a) performing multi-task sparse metric learning over the sibling child nodes under the same parent node to explicitly separate their commonly-shared metric from their node-specific metrics; and (b) propagating the node-specific metric for the parent node to its sibling child nodes (at the next level of the visual tree), so that more discriminative metrics can be learned for controlling inter-level error propagation effectively. We have evaluated our hierarchical multi-task sparse metric learning algorithm over three different image sets and the experimental results demonstrated that our hierarchical multi-task sparse metric learning algorithm can obtain better performance than the state-of-the-art algorithms on large-scale image classification.

1. Introduction

Large-scale image classification (i.e., classifying millions of images into thousands or even tens of thousands of categories) is a fundamental research issue for the communities of computer vision, machine learning and multimedia computing, and it has received extensive attentions recently [1–4]. In spite of recent significant progresses on recognizing hundreds of image categories, large-scale image classification is still a challenging task because: (1) distinguishing large numbers of image categories (i.e., thousands or even tens of thousands of image categories) is inherently more complex than distinguishing among just a few, thus the accuracy rates for the state-of-the-arts methods are usually very low; (2) the computational cost at test time is a critical issue (i.e., the computational cost for a flat approach grows linearly with the number of image categories, and it is unacceptable for large-scale image classification application); (3) some image categories may have strong inter-category visual correlations, and it does not make any sense to ignore their strong inter-category visual correlations completely and learn their inter-related classifiers independently.

One promising way to address these issues is to organize large numbers of image categories hierarchically through a tree structure by exploiting their inter-category correlations. Some previous works [2,3,5] have leveraged the semantic ontologies (taxonomies) to organize large number of image categories hierarchically according to their inter-category semantic relationships. On the other hand, it is very attractive to learn a visual hierarchy directly from large amounts of training images, and some researches [6–8] have recently been done on learning label trees to organize large numbers of image categories hierarchically according to their inter-category visual correlations. To learn a label tree, a confusion matrix for N image categories is first obtained from N one versus rest (OVR) binary classifiers. Such an approach to label tree learning may seriously suffer from two problems: (a) *huge computational cost*: it could be very expensive to learn N OVR binary classifiers independently, especially when N is very large (N is typically very large for large-scale image classification); (b) *huge sample imbalance*: for a given image category (positive class), the negative samples from other $N-1$ image categories (negative classes) heavily outnumber its positive samples, thus huge numbers of negative samples may easily mislead the process

* Corresponding author.
E-mail addresses: yuzheng.xidian@gmail.com (Y. Zheng), jfan@uncc.edu (J. Fan), zhang_ji@stu.xjtu.edu.cn (J. Zhang), xbgao@mail.xidian.edu.cn, xbgao.xidian@gmail.com (X. Gao).

http://dx.doi.org/10.1016/j.patcog.2017.01.029

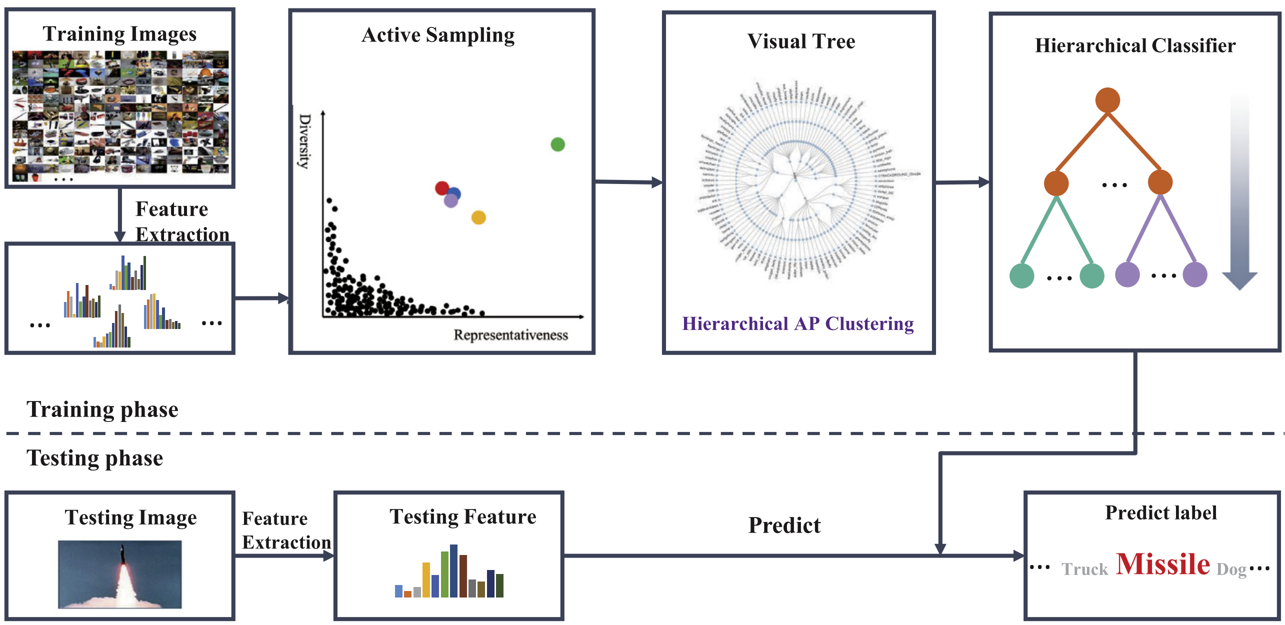

Fig. 1. The flowchart of our hierarchical multi-task sparse metric learning algorithm.

for training the OVR binary classifiers and result in incorrect OVR binary classifiers (with very low accuracy rates), which may further produce an incorrect confusion matrix for label tree construction. Therefore, other researchers have proposed visual trees learning [9,10] to organize large number of image categories hierarchically. Unlike label tree learning requires a pre-calculated confusion matrix, visual tree learning can be achieved by using the inter-category visual similarities directly, which may be able to deal with the issues of huge computational cost and huge sample imbalance effectively. However, one open problem of visual tree learning is how to provide more effective representations of image categories (which should be able to cover huge intra-category visual diversity sufficiently) and characterize the inter-category visual similarities accurately. Based on these observations, it is very attractive to develop new algorithms that are able to characterize the inter-category visual similarities accurately while reducing the computational cost dramatically.

It is also worth noting that the performance of all the algorithms for image classification crucially depends on the discrimination ability of the underlying metrics for similarity characterization. For examples, the distance metric plays an important role in the performance of *k*NN (k-nearest neighbor) classifiers, and the performance of the kernel-based algorithms (such as SVMs) can be improved significantly by adopting a proper metric to achieve more accurate similarity characterization [11–16]. Recently, many researches focus on learning large-margin metrics to achieve more accurate similarity characterization [17–20]. Even such large-margin metric learning algorithms have provided a good solution for enhancing the discrimination power of the metrics, they are not scalable for large-scale image classification application due to their huge computational complexity.

Based on these observations, a hierarchical multi-task sparse metric learning algorithm is developed to learn a tree of multi-task sparse metrics over an enhanced visual tree to achieve more accurate similarity characterization and support more effective solution to large-scale image classification. As shown in Fig. 1, our hierarchical multi-task sparse metric learning algorithm contains three key components: (a) An enhanced visual tree is first learned to organize large numbers of image categories hierarchically, and such enhanced visual tree can provide a good environment to automatically identify the inter-related tasks for multi-task sparse metric learning, e.g., the tasks for learning the large-margin metrics for the sibling child nodes under the same parent node are strongly inter-related; (b) For the root node at the first level of the visual tree, a multi-task sparse metric learning algorithm is developed to learn a multi-task sparse metric for separating its sibling child nodes more accurately; (c) For the non-root node of the visual tree, both the inter-node visual correlations (among the sibling non-root nodes under the same parent node) and the inter-level visual correlations (between the parent node and its sibling child nodes at the next level of the visual tree) are leveraged to learn a tree of multi-task sparse metrics, so that more discriminative metrics can be learned for controlling inter-level error propagation effectively.

The rest of the paper is organized as follows. A brief review of related work is presented in Section 2. In Section 3, we introduce our work on constructing the enhanced visual tree. In Section 4, we present our work on hierarchical learning of a tree of multi-task sparse metrics over the visual tree. Section 5 demonstrates our experimental results, followed by conclusions.

2. Related work

Recently, metric learning [15–17,19,21–23] has received extensive attentions in communities of computer vision, machine learning and multimedia computing, especially Mahalanobis metric learning. Weinberger et al. have developed a metric learning approach to support nearest neighbor classification (LMNN) [19], which aims to learn a Mahanalobis distance by forcing the samples from the same class to be closer while the samples from different classes are far away. Parameswaran et al. have extended such LMNN approach to a multi-task setting and a multi-task metric learning approach [24] is proposed. The multi-task metric learning approach aims to learn one common metric shared among multiple inter-related tasks and multiple task-specific metrics simultaneously. By separating the commonly-shared metric from the task-specific metrics explicitly, the multi-task metric learning approach can obtain more discriminative metrics and achieve higher classification accuracy rates.