Signal Processing 112 (2015) 137–145

Contents lists available at ScienceDirect

Signal Processing

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

Large-scale multi-task image labeling with adaptive relevance discovery and feature hashing

Cheng Deng [a], Xianglong Liu [b,*], Yadong Mu [c], Jie Li [a]

[a] School of Electronic Engineering, Xidian University, Xi'an 710071, China
[b] State Key Lab of SDE, Beihang University, Beijing 100191, China
[c] AT&T Labs Research, 200 S Laurel Ave, Middletown, NJ 07748, USA

ARTICLE INFO

Article history:
Received 18 March 2014
Received in revised form
23 June 2014
Accepted 23 July 2014
Available online 21 August 2014

Keywords:
Image classification
Multiple tasks
Feature hashing
Relevance discovery

ABSTRACT

It remains challenging to train an effective classifier for the new image classification tasks provided with only a few or even no labeled samples. Although multi-task learning approaches have been introduced into this field to exploit available label information to boost classification accuracy, these approaches discover intrinsic task relationships only at task level, which will lead to limited useful labels being exploited and shared. Motivated by clustered multi-task learning, this paper proposes a robust multi-task feature hashing learning algorithm for image classification. Specifically, the original input samples are first projected into a low-dimensional hash feature subspace, upon which not only the inherent relatedness but also the fine-grained clustering among samples can be revealed well. Then, the task relationships are captured by interacting at task level as well as at feature level, and finally the auxiliary labels can be shared across different tasks. We conduct extensive experiments on three large-scale multi-label image classification datasets, and results demonstrate the superiorities of the proposed formulation in comparison with several state-of-the-arts.

1. Introduction

The past decade has witnessed the tremendous accumulation of digitalized images and videos in broadcasting archives or web-sharing sites. The data proliferation motivates the development of high-performance content-based indexing algorithm in order to facilitate real-time retrieval operation. The multimedia data are often associated with semantic side information (e.g., tags, labels, surrounding text, or user-provided annotations), which enables learning the visual semantic models from labeled training data. However, conventional machine learning methods often only work well given sufficient training data. In real-world problem settings, obtaining meticulous labeling is often labor-intensive and expensive. The scarcity of labeled data is a severe challenge in many tasks [1,2], and much research endeavor has been devoted to overcome it. For example, semi-supervised learning (SSL) methods [3] utilize the cluster assumption and manifold structure underlying the data to regularize the supervised learning models. However, SSL is reported to produce limited performance in real-world problems [4].

This paper seeks a new approach for visual image classification. Under a unified formulation, it addresses both the above-mentioned data scarcity issue and the scalability of the learned models in large-scale regime.

First, recent achievement on multi-task learning (MTL) sheds light on learning a new task with limited samples by harnessing other inherently related tasks. In the context of

* Corresponding author at: Room G611, New Main Building, Beihang University, Xueyuan Rd, Haidian, 100191, Beijing.
E-mail address: xlliu@nlsde.buaa.edu.cn (X. Liu).
URL: http://www.nlsde.buaa.edu.cn/~xlliu (X. Liu).

http://dx.doi.org/10.1016/j.sigpro.2014.07.017
0165-1684/

image classification, different labels are often semantically or statistically related. Treating the pursuit of the model for specific label as a unique task, it is a nature idea to transfer useful information across relevant tasks in various forms, such that data scarcity issue can be largely mitigated. In traditional MTL methods [5–8], all tasks are surmised to be related to each other, which simplifies the numerical optimization yet fails to fully capture the problem structure in real-world applications. Rather than blindly assuming a full-correlation model, some recent works on MTL explore the ideas of discovering the cluster structure at task level [9], or identifying relevant and irrelevant tasks [10]. Inspired by the work of [11], we go one step further that we explore the inherent relationship both at task level and at feature level.

Second, the rich information in visual contents demands high-dimensional feature representations, yet parsimony is an important criterion for storing large-scale data (e.g., the samples may be in million or even billion order). Fortunately, most of the feature dimensions are often weakly relevant to current tasks or redundant to each other. To ensure the final model to be scalable, we therefore incorporate dimension reduction into the formulation. Specifically, we generate a set of binary hashing functions which project the original high-dimensional data into a more compact binary feature space. With deliberate regularization on such projection, we enjoy enhanced scalability while still discerning the inherent relationship among different tasks.

Putting above-mentioned considerations together, we propose the RMTHL (robust multi-task feature hashing learning) algorithm for accomplishing adaptive relevancy discovery in multi-task setting, by taking full exploitation to the task-level and feature-level relations. It also seamlessly incorporates the idea of hashing-based feature dimension reduction, which largely reduces the data storage overhead (only compact hash code is required to be stored for each sample) and more efficient classification models (they are defined on the hash codes rather than original features). We verify our algorithm on three popular image classification datasets, and experimental results demonstrate that our proposed RMTHL has good accuracy and scalability, and is superior to some representative methods.

The paper is organized as follows. Section 2 presents related work on multi-task learning and feature hashing. In Section 3, we detail the description and formulation of the algorithm and its optimization technique. Our experiments and analysis are depicted in Section 4. Finally, Section 5 concludes our paper.

2. Related work

In this section, we briefly introduce two lines of research works which are tightly relevant to our proposed model, e.g., multi-task learning and feature hashing.

2.1. Multi-task learning

MTL has been applied successfully in many applications including object recognition [5], handwritten digit recognition [12], and natural language processing [13].

Currently, many efforts have been made to devise various MTL algorithms from different perspectives. Zhang et al. [8] learned the model parameters and the task relationship simultaneously, and shared a common prior in a Bayesian model. Some MTL algorithms with a composite regularization have been presented to capture different types of relationships using regularization, such as ℓ_1-norm [14], $\ell_{2,1}$-norm [15], and trace-norm [16]. Kim et al. [17] modeled the relatedness of multiple tasks using structured sparsity penalties called tree-guided group lasso. Chen et al. [18] proposed a linear MTL formulation where the model parameter can be decomposed into a sparse component and a low-rank component. Some approaches have been proposed to discover the underlying task clusters or group structure. For example, Zhou et al. [9] assumed that multiple tasks follow a clustered structure, and shown that this approach is equivalent to alternating structure optimization [19] which assumed the tasks share a low-dimensional structure. Gong et al. [10] simultaneously captures a common set of features among relevant tasks and identifies outlier tasks. Kang et al. [20] modeled the task relatedness as learning shared features among the tasks.

2.2. Feature hashing

Hashing is an important method to achieve near neighbor fast similarity search in sub-linear time by mapping highly similar data together, which is widely used in post-estimation [21], image retrieval [22,23], and similarity learning [24]. Up to now, many hashing methods have been proposed, such as locality-sensitive hashing (LSH) [25], semantic hashing [26], spectral hashing [27], and so on. LSH approaches construct a family of hashing functions via linear projection over random directions. Data-dependent hashing methods have drawn extensive attractions recently, where compact hash codes are learned. Semantic hashing is proposed based on stacked restricted Boltzman machine. Spectral hashing utilizes spectral graph partitioning in the metric learning phrase to get compact hash codes of training data. Since spectral hashing takes a restrictive and unrealistic assumption that data are uniformly distributed in a hyper-rectangle, several sequential new methods relax this restrictive assumption, such as shift-invariant kernel hashing [28,29]. Current hash function learning methods focus on hashing in several special and important settings, for example, multi-dimensional spectral hashing [30].

3. The proposed algorithm

3.1. Notations and overview

In multi-label image classification, a label prediction model is needed for each specific image label. We assume that for each label, a number of annotated samples are provided as positive samples, otherwise an image is treated as negative one. In the terminology of multi-task learning, learning the model for specific label defines a unique task. Suppose there are T image labels (correspondingly the same number of tasks) in total. For notation simplicity, we stack the samples of the same label together