Signal Processing: Image Communication 64 (2018) 59–67

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

Multi-task classification with sequential instances and tasks

Wei Xu [a,b,*], Wei Liu [a], Haoyuan Chi [a], Xiaolin Huang [a], Jie Yang [a,**]

[a] Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China
[b] Lab for Media Search, School of Computing, National University of Singapore, Singapore

ARTICLE INFO

Keywords:
Classification
Multi-task learning
Curriculum learning
Self-paced learning

ABSTRACT

In this paper, we propose a novel multi-task classification framework, called Multi-Task classification with Sequential Instances and Tasks (MTSIT). Different from previous works, which treat all tasks and instances equally, MTSIT is inspired by the cognitive process of human brain that often learns from easier tasks to harder tasks. Specifically, the method attempts to jointly learn the task curriculum (learning order of tasks) and the instance curriculum (learning order of instances) by introducing a self-paced item for the instances of each task in the existing multi-task learning framework Sequential Multi-Task learning (SeqMT), which transfers information from the previously learned tasks to the next ones through shared task parameters. To effectively solve MTSIT, we also propose an optimization algorithm in which the instance curriculum and the task curriculum alternate between two paradigms, Tasks-to-Instances and Instances-to-Tasks (TIIT). In the tasks-to-instances step, the learner conducts the instance curriculum when the task curriculum has been fixed, while in the instances-to-tasks step, the task curriculum is learned when the instance curriculum in each task has been settled down. Our TIIT method is based on an error bound of the proposed MTSIT. Experimental results on three real world datasets demonstrate the effectiveness of our method.

1. Introduction

Multi-Task learning (MTL) [1] is the learning framework where several tasks are learnt jointly. While traditional machine learning algorithms generally learn each task independently. They often require a large mount of labeled data to achieve satisfying performance. However, in real world it is time consuming and expensive to get large amounts of labeled data. To handle this problem, many researchers tend to explore shared knowledge among related tasks. This approach has been proved experimentally effective [2].

In this paper, our focus is the model parameter based multi-task learning method. This kind of method generally assumes that related tasks share the similar model parameter values. Most of previous MTL models rest on the assumption that during the training process, all the tasks and all the instances per task should be treated equally. But this assumption may not be optimal in real scenarios. In real world applications, there may be groups of tasks with little similarity between tasks from different groups. To alleviate this issue, Pentina proposed a curriculum learning method [3] for multiple tasks to find the best learning order of tasks. It is named as Sequential Multi-Task learning (SeqMT) [4]. Here the learning order of tasks is called task curriculum. With the learned task curriculum, SeqMT processes multiple tasks sequentially to transfer information between sequential tasks. Concretely, it regularizes the model parameters of previous learned tasks and the next ones to be similar to each other.

In SeqMT, the curriculum is solely learned for tasks. In this paper, we show that instance curriculum should also be exploited in multi-task learning. Generally, a reasonable curriculum for students not only requires to have related tasks to fit their learning abilities but also includes examples of suitable easiness in each task to match their learning paces. Likewise, learning from both the task curriculum and the instance curriculum is better than learning from only one of them.

The Instance curriculum can be learned by using Self-Paced learning (SPL) [5]. SPL is based on the idea that data from easy ones to more complex ones are gradually involved in learning. Moreover, it has been empirically proved to be effective in the task with heavy noise [5,6].

We propose a new multi-task classification paradigm that introduces a self-paced item for the instances of each task as the instance curriculum into the SeqMT framework. It is hence named as Multi-Task classification with Sequential Instances and Tasks (MTSIT). By considering both the task curriculum and the instance curriculum in training, MTSIT can

* Corresponding author at: Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, China.
** Corresponding author.
E-mail addresses: xuronaweizida@gmail.com (W. Xu), liuwei.1989@sjtu.edu.cn (W. Liu), chihaoyuan6@sjtu.edu.cn (H. Chi), xiaolinhuang@sjtu.edu.cn (X. Huang), jieyang@sjtu.edu.cn (J. Yang).

https://doi.org/10.1016/j.image.2018.02.013
Received 6 December 2017; Received in revised form 23 February 2018; Accepted 23 February 2018
Available online 27 February 2018

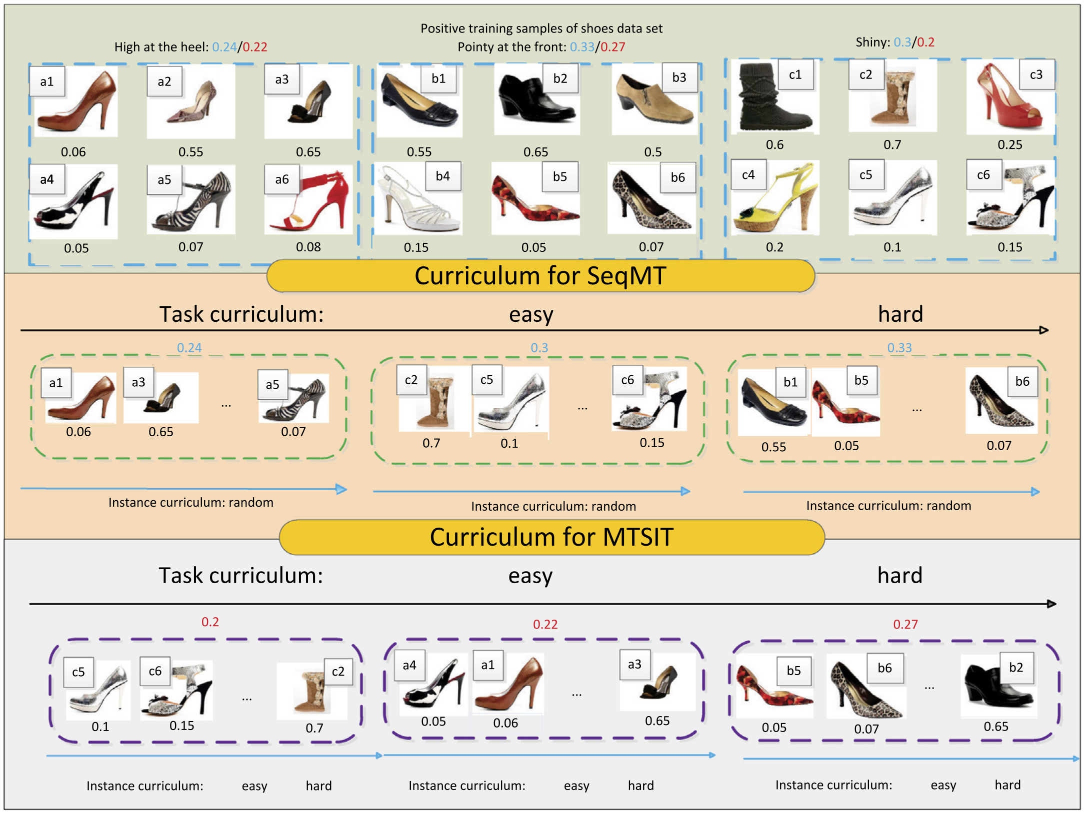

Fig. 1. Comparison of SeqMT and MTSIT on attribute recognition using data in Shoes augmented with attributes data set [7]. The black value represents the training loss of the instance and the blue and red value represent the training loss of the first iteration learned by SeqMT and MTSIT respectively. SeqMT tends to select the random samples in each task and the easiest task. MTSIT inclines to select the easiest sample in each task and the associated updated version easiest task.

learn better model parameters. For example, Fig. 1 shows some data on a real dataset, named Shoes augmented with attributes [7]. Three tasks are employed for demonstration. The black value under the instance represents the training loss value of the instance. Hence, if the training loss of a instance is small means this instance is easy. The complexity of tasks also exists. The colored value behind the task represents the training loss value of the task and a smaller training loss corresponds to an easier task. The blue and red value represent the task training loss of the first iteration learned by SeqMT and SeqMT respectively. Training with the above data, SeqMT may lead to overfitting to a data subset due to hard instances in each task while ignoring its easy samples. For example, in Fig. 1, as many of the chosen samples of task "pointy to the front" by SeqMT are very difficult for classification and all the samples in this task are considered equally during the training process, SeqMT will easily overfit to those difficult samples in the training dataset. In contrast, MTSIT, considering both the task curriculum and the instance curriculum, produces a curriculum that reasonably finds the relationship between tasks and instances of them. The two layer curriculum can quickly get the easy and related information, hence better model parameters are able to be learned. This hypothesis is validated in the experimental section.

As the curriculum may significantly affect the overall performance, we study the task curriculum and the instance curriculum by PAC-Bayesian theory [8]. Specifically, we adopt PAC-Bayesian theory to provide an error bound to quantify both the effectiveness of the task curriculum and that of the instance curriculum. Hence we propose our theoretical bound based algorithm dubbed Tasks-to-Instances and Instances-to-Tasks (TIIT) to learn a favorable sequence for training.

There are three main contributions in this paper: (1) We propose a new multi-task classification framework MTSIT that considers both the complexity of tasks and that of instances, and this framework is able to be applied to various problems. (2) An effective optimization algorithm TIIT is developed for the model. (3) We provide a theoretical explanation for MTSIT, which is the first to incorporate SPL into the multi-task classification based on the PAC-Bayesian theory [8]. Experimental results on three real-world datasets demonstrate the effectiveness of the proposed approach.

2. Related work

2.1. Multi-task learning

That there exists shared knowledge among related tasks is one of the basic assumptions in MTL. Algorithms based on this idea have been proved experimentally more effective than those methods that train each individual task independently [1]. Many multi-task learning methods have been proposed, which can be categorized into two groups [9,10]. The first group assumes that there exists a common yet low-rank feature representation shared by all the tasks [11–14]. The other group assumes that related tasks share the similar model parameter values [15–17]. To combine the two classes of approaches, some works jointly learn feature representation and model parameter relation in a unified framework [18,19]. Recently, some works exploit the complexity of tasks. For example, SeqMT [4] uses a heuristic algorithm to rank all the tasks and learns the tasks sequentially for multi-task classification. Different from the existing multi-task learning methods, our proposed MTSIT for multi-task learning aims to learn a robust classification model by exploiting both the complexity of instances and that of tasks.

2.2. Self-paced learning

Inspired by the learning process of humans, self-paced learning [5] is proposed. It is based on the idea that training data from easy ones to more complex ones are gradually involved in learning. It has