Pattern Recognition 48 (2015) 3249–3257

Contents lists available at ScienceDirect

Pattern Recognition

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

Multi-task proximal support vector machine ☆

Ya Li [a,1], Xinmei Tian [a,*], Mingli Song [b], Dacheng Tao [c]

[a] University of Science and Technology of China, Hefei, Anhui 230026, PR China
[b] Zhejiang University, Hangzhou, Zhejiang, China
[c] University of Technology, Sydney, Australia

ARTICLE INFO

Article history:
Received 17 October 2014
Received in revised form
15 January 2015
Accepted 16 January 2015
Available online 3 February 2015

Keywords:
Multi-task learning
Support vector machines
Proximal classifiers

ABSTRACT

With the explosive growth of the use of imagery, visual recognition plays an important role in many applications and attracts increasing research attention. Given several related tasks, single-task learning learns each task separately and ignores the relationships among these tasks. Different from single-task learning, multi-task learning can explore more information to learn all tasks jointly by using relationships among these tasks. In this paper, we propose a novel multi-task learning model based on the proximal support vector machine. The proximal support vector machine uses the large-margin idea as does the standard support vector machines but with looser constraints and much lower computational cost. Our multi-task proximal support vector machine inherits the merits of the proximal support vector machine and achieves better performance compared with other popular multi-task learning models. Experiments are conducted on several multi-task learning datasets, including two classification datasets and one regression dataset. All results demonstrate the effectiveness and efficiency of our proposed multi-task proximal support vector machine.

1. Introduction

Given the explosive growth the use of imagery in the era of big data, visual recognition has become an important problem. Various image classification and recognition methods have been proposed and have achieved much success [1–9]. Some feature learning methods are also proposed to improve the performance of image classification and recognition [10–13]. When learning a visual recognition task, it can often be viewed as a combination of multiple correlated subtasks [14]. Considering multi-label image classification, for example, one particular image may contain multiple objects corresponding to different labels. Obviously, there are correlations among these labels. Traditional single-task learning methods, for example, SVMs and Bayesian models, learn to classify these labels separately and ignore correlations among them. It would be desirable to explore shared information across different subtasks and apply the information to learn all the subtasks jointly. Inspired by this idea, various methods are proposed to learn multiple tasks jointly rather than separately. This is often called the multi-task learning (MTL) [15], learning to learn [16] or inductive bias learning [17]. All these methods tend to learn multiple tasks together and improve the performance of single-task learning models.

The most important and difficult problem in multi-task learning is to discover the shared information among tasks and maintain the independence of each task. Considering the classification of vehicles (see Fig. 1), we have various types of vehicles, such as sports cars, family cars and buses corresponding to different classification tasks. These cars have shared features as well as unique characteristics. For example, all cars have four wheels and two headlights. However, sports cars usually have a lower and racing body, family cars often have medium size, and buses have a bigger body. Single-task learning only uses the information of the independent task, while multi-task learning will use all the information among the tasks. If a multi-task learning method can find the shared features of these vehicles and distinguish differences among the vehicles, each learning task will have much more additional information from other tasks. Conversely, noise will be added to the current learning task.

Existing multi-task learning methods mainly have two ways to discover relationships among different tasks. One way is to assume that different tasks share common parameters [18,14,19–23] such as a Bayesian model sharing a common prior [14] or a

☆This work is supported by the NSFC under the Contract nos. 61201413 and 61390514, the Fundamental Research Funds for the Central Universities Nos. WK2100060011 and WK2100100021, and the Specialized Research Fund for the Doctoral Program of Higher Education No. WJ2100060003. Australian Research Council Projects: DP-140102164, ARC FT-130101457, and ARC LP-140100569.

* Corresponding author. Tel.: +86 183 551 026 90; fax: +86 551 636 013 40
E-mail addresses: muziyiye@mail.ustc.edu.cn (Y. Li), xinmei@ustc.edu.cn (X. Tian), brooksong@ieee.org (M. Song), Dacheng.Tao@uts.edu.au (D. Tao).
[1] CAS Key Laboratory of Technology in Geo-spatial Information Processing and Application System, University of Science and Technology of China, China

http://dx.doi.org/10.1016/j.patcog.2015.01.014
0031-3203/

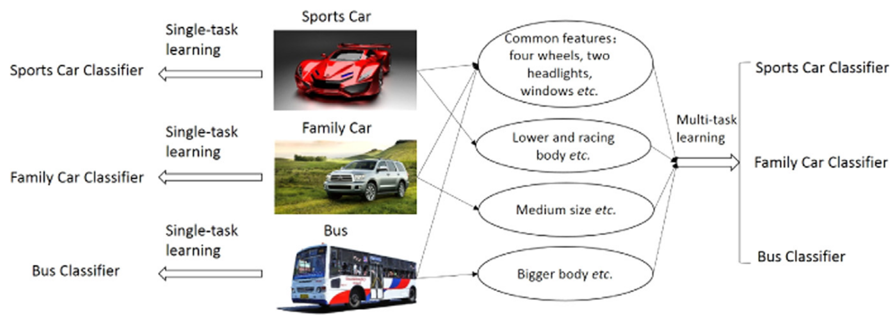

Fig. 1. An example of single-task learning comparing with multi-task learning.

large-margin model sharing a mean hyperplane [19]. The other way to learn the relatedness is to find latent feature representation among these tasks [24–26], for example, learning a sparse representation shared across tasks [25]. Existing multi-task learning methods mainly have two defects. First, some multi-task learning models have a complicated theoretical foundation, which leads to implementation difficulties. For example, a nonparametric Bayesian model usually has many assumptions and many parameters to select. Second, the efficiency is low, especially when the dataset has a large number of data points and a high dimensional feature. Our goal is to find an easily implemented multi-task learning method with high efficiency and comparable performance. In this paper, we propose a multi-task learning method based on the proximal support vector machine (PSVM) [27] and apply it to two classification datasets and one regression dataset. PSVM was proposed by Fung and Mangasarian and is different from the standard SVM [28]. PSVM also utilizes the large margin idea by assigning the data points to the closest of two disjoint hyperplanes, which are separated as far as possible. However, PSVM has looser constraints than does standard SVM, with comparable performance and much lower computational cost. Inspired by the idea of PSVM and the advantages of multi-task learning, we derive a multi-task proximal support vector machine (MTPSVM). All data examples of all tasks are needed to learn MTPSVM simultaneously. It will absolutely slow the computing process if the dataset is a large-scale one. In this paper, we develop a method to optimize the procedure of learning MTPSVM that greatly improves efficiency. Based on the idea of PSVM for unbalanced data, we also apply this to MTPSVM. Finally, we propose proximal support vector regression for regression problems, which is not discussed in PSVM [27], and extend it to multi-task problems.

MTPSVM has two primary merits compared with other multi-task learning methods. First, MTPSVM is easily implemented by just solving a quadratic optimization problem with equality constraints. Second, MTPSVM has much lower computational cost and can be applied to a large-scale dataset. We will demonstrate that the computational time of MTPSVM relies primarily on the feature dimension of the data rather than on the number of data points.

We organize the remainder of this paper as follows. Section 2 reviews previous works in multi-task learning. In Section 3, we first briefly introduce the proximal support vector machine and then give a specific derivation of the proposed multi-task proximal support vector machine. The derivation of multi-task proximal support vector regression will be presented in Section 4. In Section 5, experiments on several datasets are presented. Section 6 presents our study's conclusions.

2. Related work

Multi-task learning has been proven more effective than single-task learning by many works via both theory analysis and extensive experiments. For example, Baxter proposed a novel model of inductive bias learning to learn multiple tasks together and derived explicit bounds which demonstrated that multi-task learning gave better generalization than single-task learning [17]. Another work conducted by Ben-David and Schuller developed a useful notion of task relatedness and better generalization of error bounds for learning multiple related tasks based on one special type of relatedness of tasks [29]. Both studies prove the merits of multi-task learning in theory. Various experiments also demonstrate that multi-task learning can achieve better performance than can single-task learning, e.g., experiments on School Dataset [19,30,25,31], Landmine Dataset [14,24]. Multi-task learning can achieve much better performance than single-task learning especially when the amount of training data is limited.

Due to the effectiveness of multi-task learning, many single-task learning methods are extended to multi-task learning ones, such as neural networks, nearest neighbor learners, Bayesian model and SVM. For example, multi-task learning methods are implemented by sharing hidden nodes in neural networks or using nearest neighbor learners [15,32]. Bayesian is another popular model for multitask learning. It assumes dependencies between various models and tasks [33,34]. Models can be learned by hierarchical Bayesian inference with shared parameters treated as hyperparameters at a higher level than the single-task model parameters. In recent years, nonparametric Bayesian models and infinite latent subspace learning have become popular in multi-task learning. Rai and Daume proposed an infinite latent feature model to automatically infer the dimensionality of the task subspace. They learned a multi-task learning model using the Indian Buffet Process as the nonparametric Bayesian prior [18]. Consider the success of SVM in single-task learning, support vector machines are popular in multi-task learning. Many multi-task learning methods are developed based on support vector machines with different assumptions or priors [35,19,30,24]. An infinite latent SVM for multi-task learning is derived using nonparametric Bayesian models with regularization on the desired posterior distributions [35]. Evgeniou and Pontil proposed a novel multi-task learning method based on the minimization of regularization functions, similar to support vector machines [19]. Based on the work of [19], a more specific and general derivation of kernel method was developed in [30]. Jebara proposed a maximum entropy discrimination method for multi-task learning based on the large-margin support vector machines [24]. It gives extensions of feature selection and kernel selection for multi-task learning. The idea of our multi-task learning method is similar to [19]. The difference is that our multi-task learning method is based on proximal support vector machine rather than on the standard support vector machines. This results in an easier implementation and lower computational cost.

As mentioned above, learning latent common features across tasks and sharing common parameters are two important ways to model the relatedness of multi-task learning. For learning latent common features, a framework was proposed to learn sparse