Neurocomputing 123 (2014) 100–109

Contents lists available at ScienceDirect

Neurocomputing

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

Online heterogeneous feature fusion machines for visual recognition

Shuangping Huang [a], Lianwen Jin [b,*], Yuan Fang [c], Xiaoxin Wei [b]

[a] College of Engineering, South China Agricultural University, Guangzhou, PR China
[b] School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, PR China
[c] College of Computer & Information Science, Northeastern University, Boston, USA

ARTICLE INFO

Article history:
Received 25 November 2012
Received in revised form
13 April 2013
Accepted 19 June 2013
Communicated by D.S. Huang
Available online 2 July 2013

Keywords:
Online optimization
Heterogeneous feature fusion
Visual recognition
Group LASSO
Multiple kernels learning

ABSTRACT

Heterogeneous Feature Fusion Machines (HFFM) is a kernel based logistic regression model that effectively fuses multiple features for visual recognition tasks. However, the batch mode solution for HFFM, 'Block Coordinate Gradient Descent' (BCGD) has the same low efficiency and poor scalability as the most batch algorithms do. In this paper, we describe a newly developed online learning algorithm in multiple Reproducing Kernel Hilbert Spaces for solving HFFM model. This new algorithm is called OLHFFM, i.e. Online HFFM. OLHFFM is novel combination of kernel-based learning technique with dual averaging gradient descent methods. In addition, group LASSO regularization technique is used in OLHFFM for finding important explanatory coefficients that are related to support samples in group manner. The effectiveness of OLHFFM has been demonstrated by a number of experiments that were conducted on public event, object dataset, as well as on large scale handwritten digital dataset. Using the OLHFFM approach, we have achieved almost equivalent recognition performance to that using batch-mode approach. Experiments conducted on both MIT Caltech-6 and challenging VOC2011 TrainVal object datasets show that OLHFFM is superior in performance to kernel based online learning approaches such as ILK or NORMA. In addition, the classification performance of OLHFFM approach as demonstrated by the experiments conducted on large scale MNIST dataset is comparable to or better than that of the current state-of-the-art online multiple kernel learning approaches such as OM-2, UFO-MKL, OMCL and OMKL. Extensive experiments on visual data classification demonstrate the effectiveness and robustness of the new OLHFFM approach.

1. Introduction

Impressive progress has been made in visual recognition field recently [1–25,38,41,44,47]. An important part of recent work has been focusing on a range of advanced image descriptors, for example, SIFT (Scale Invariant Feature Transform) [2], GIST [3], Histogram of Oriented Gradients [4], Local Binary Pattern (LBP) [5], CTM (Color Texture moment) [6], shape context [7] and so on. Different features describe different aspects of the visual characteristics and cover different visual recognition cues such as appearance, texture, shape and color. They are complementary to each other. For more and more complex visual recognition tasks, one cannot just use single type feature since it does not provide enough discriminative information. For this reason, combination of heterogeneous features is gaining popularity lately for more complex visual recognition tasks [8–21,47].

Technical challenges still exist for fusing multiple features in general way. A large number of publications can be found in this subject, even though they could be under different names such as "multiple view learning" or "multiple features/heterogeneous feature fusion". These publications fall into the following categories according to the intrinsic nature of their formulations: (1) multiple feature fusion by projection or subspace learning [9,10]; (2) multiple feature fusion by combining different kernels that corresponds to different measures of similarity for different representations [11–15]; (3) multiple feature fusion by means of spectral embedding, e.g. multiview (MSE) [16] or distributed (DSE) [17]; and (4) methods based on data graph wherein convexly combining the graph Laplacians on different views [18,19]. Besides, there are some case-by-case algorithmic instantiation in solving specific real-world problems, for example, Co-LapSVM and Co-LapRLS within co-regularization framework for semi-supervised learning [20]; m-SNE based on stochastic neighbor embedding [21] etc. Multiple Kernel Learning (MKL) based models are among the most popular ones because it is the most reasonable approach for combining multiple information sources. MKL combines different kernels for different features by a weighted summation. The weight for each feature does not depend on one certain sample

* Corresponding author. Tel.: +86 20 87113540.
E-mail addresses: shuangping.huang@gmail.com (S. Huang), lianwen.jin@gmail.com, eelwjin@scut.edu.cn (L. Jin).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.06.004

and remains the same across all the samples [24]. Recently, a new HFFM model gains popularity. In this new HFFM model, the weights of kernels vary from sample to sample. This leads to nonlinear fusion of the multiple kernels functions. It makes kernel combination more feasible and thus promotes its fusion capability [24,47].

For visual recognition tasks, batch mode solution has been used for heterogeneous feature fusion. It is well known that batch solution approach has the limit of poor scalability, low efficiency, and high cost [24,35]. It even becomes impractical to use batch solution approach when one has to handle millions of image samples. As a result, online learning algorithms have gained popularity for their high efficiencies in large-scale data analysis [26-33,40,50]. Another advantage of online algorithm is the ability to "include human in the loop" with robotic vision.

In this paper, we describe a novel online algorithm called OLHFFM in multiple Reproducing Kernel Hilbert Spaces that combines group LASSO sparse method and dual averaging sub-gradient learning technique. This online algorithm is used to solve HFFM model efficiently and it can be used for a wide range of visual recognition tasks such as event recognition, object categorization and so on. Different than standard online MKL, the solution of HFFM tends to depend on a subset of low-noise samples. Group LASSO is used to select explanatory samples and remove noisy samples in HFFM model for the classifying function. In our work we demonstrated the feasibility of implementing non-linear multiple kernel fusion technique in an online mode.

The remainder of this paper is organized as follows. Section 2 reviews the related online learning works especially that focuses on online solutions for standard MKL model. Section 3 states the problem of HFFM models and formulizes the online HFFM model solution; Section 3 presents some analytical and comparative experiments on a variety of visual recognition task delivered on some publicly available benchmark datasets and show some findings. Section 4 concludes this study with future work.

2. Related work

Some examples of online algorithms that are used in linearly separable cases include Rosenblatt's Perceptron [26], FOBOS method developed by Duchi and Singer [27], RDA (Regularized Dual Averaging) [28,50] and DA-GL (Dual Averaging-Group LASSO) [29]. Among these, FOBOS can be considered as a general framework for stochastic gradient with arbitrary regularization. It alleviates the problems of non-differentiability in cases such as ℓ_1-regularization by taking analytical minimization steps interleaved with sub-gradient steps. RDA (Regularized Dual Averaging) [28] method is proposed by Lin Xiao for solving regularized learning problems under online setting as FOBOS does. The uniqueness of RDA is that it updates model parameter vector by solving a simple optimization problem on each round that involves average of all the past sub-gradients of the loss functions and the whole regularization term. That is the main difference between RDA and FOBOS. In essence, RDA adjusts the learning variables not only using information from the single coming example but also from sub-gradients of loss functions as to past examples. However, only simple LASSO regularization is considered in RDA for sparsity. DA-GL [29] inherits the idea of RDA and it is extended for solving a group LASSO regularized optimization problem in the original feature space. Moreover, various variants of group LASSO such as sparse group LASSO [21], group LASSO with overlap and graph LASSO [22] can be adopted as online algorithms' regularization item for finding important explanatory variable group. It is noted that attributes in the feature space form a group here. That is to say, d-dimension feature vector is divided into G groups with d_g (the number of attributes in g-th group). The number d_g is usually assumed greater than 1. When a group is sparsified by means of various group LASSO methods, corresponding feature attributes are set to zero in the model.

For linearly inseparable data analysis another family of online algorithms with kernel integration is used. Some examples of single-kernel based online algorithms include NORMA (Naïve Online R-reg Minimization Algorithm) [30], ILK (Implicit online Learning with Kernels) [31] and so on. Both NORMA and ILK perform gradient descent with respect to ℓ_2-regularized instantaneous risk in Reproducing Kernel Hilbert Spaces (RKHS). The main difference between them is that NORMA implements explicit parameter updates and ILK implements implicit ones. However, both of them yield no sparse solutions since they involves ℓ_2-norm in RKHS which produces only soft shrinkage on hypothesis. In practice, it is often desirable to seek LASSO [18] or group LASSO [19] with sparsity at group level or individual level. Some state-of-the-art multiple kernel based online algorithm examples including UFO-MKL [14], OBSCURE [33], and OM-2 [11] adopt feasible regularization technique to obtain tunable sparsity or make optimization problem easier. Among these, UFO-MKL mixes elements of group p-norm and LASSO, i.e. forming elastic net kind of regularization which separately provides an easy optimization problem and induces feasible levels of sparsity in the domain of the kernels. Stochastic gradient descent and mirror descent are combined and used to solve this elastic net kind regularized MKL problem. The solution of the optimization problem will lead to a selection a subset of the F kernels. OBSCURE is proposed to solve p-norm version of the standard MKL model. It highlights two-stage optimization method. The first stage is an online initialization procedure that determines quickly the region of the space where the optimal solution lives. The second one refines the solution found by the first stage. OM-2 uses the 'Follow the Regularized Leader' [48,49] framework to solve group p-norm regularized MKL problem. The other two algorithms including OMKL [15] and OMCL [32] focus not on regularization but on decomposing MKL solution into two separate tasks. OMKL uses deterministic or stochastic approaches to combine binary predictions or real-valued outputs from multiple kernel classifiers. The deterministic approach updates all kernel classifiers for every misclassified example, while the stochastic approach chooses a classifier(s) randomly for updating according to some sampling strategies. Different setup, i.e. deterministic or stochastic, binary predictions or real-valued outputs, forms OMKL series algorithms. OMCL is a wrapper algorithm using a two-layer structure, which can use most of the known online learning methods as base algorithms. However, all of these algorithms are part of the standard Multiple Kernel Learning (MKL) family. That means that they aim to obtain multiple kernels classifier and their linear combinations from a pool of given kernels in an online fashion. Moreover, the weights *wm* for the *mth* kernel remain the same across all the samples. We emphasize that although a number of approaches have been proposed to solve the optimization problem related to MKL, little work has been done to address online HFFM learning. To the best of our knowledge, this is the first theoretic study that addresses the online HFFM problem.

3. Online HFFM algorithm

3.1. Preliminaries

Before presenting OLHFFM learning method, we first describe briefly HFFM [24,47] and introduce some basic notations for classification. Assume $\{\boldsymbol{x_i}, y_i\}$ is an input–output pair. Here $y_i \in \{1, 0\}$ is the label of a sample for binary classification problem and