Signal Processing 110 (2015) 101–107

Contents lists available at ScienceDirect

Signal Processing

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

Multiview Hessian regularized logistic regression for action recognition

Weifeng Liu [a], Hongli Liu [a], Dapeng Tao [b,c,*], Yanjiang Wang [a], Ke Lu [d]

[a] China University of Petroleum (East China), PR China
[b] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, P.R. China
[c] The Chinese University of Hong Kong, Hong Kong, P.R. China
[d] University of the Chinese Academy of Sciences, P.R. China

ARTICLE INFO

Article history:
Received 6 March 2014
Received in revised form
28 May 2014
Accepted 2 August 2014
Available online 11 August 2014

Keywords:
Hessian
Logistic regression
Action recognition
Multiview learning
Manifold learning
Semi-supervised learning

ABSTRACT

With the rapid development of social media sharing, people often need to manage the growing volume of multimedia data such as large scale video classification and annotation, especially to organize those videos containing human activities. Recently, manifold regularized semi-supervised learning (SSL), which explores the intrinsic data probability distribution and then improves the generalization ability with only a small number of labeled data, has emerged as a promising paradigm for semiautomatic video classification. In addition, human action videos often have multi-modal content and different representations. To tackle the above problems, in this paper we propose multiview Hessian regularized logistic regression (mHLR) for human action recognition. Compared with existing work, the advantages of mHLR lie in three folds: (1) mHLR combines multiple Hessian regularization, each of which obtained from a particular representation of instance, to leverage the exploring of local geometry; (2) mHLR naturally handles multi-view instances with multiple representations; (3) mHLR employs a smooth loss function and then can be effectively optimized. We carefully conduct extensive experiments on the unstructured social activity attribute (USAA) dataset and the experimental results demonstrate the effectiveness of the proposed multiview Hessian regularized logistic regression for human action recognition.

1. Introduction

The prodigious development of devices capable of digital media capture and internet technology makes vast volumes of multimedia data be uploaded and shared on the internet. It is an immediate need for effective techniques that can help manage this massive amount of data including video annotation, multimedia retrieval, multimedia understanding [51,52]. Due to the expensive labor of labeling a large number of media data for model learning, semi-supervised learning (SSL) has attracted much attention since it can improve the generalization ability of a learning model by exploiting both a small number of labeled data and a large number of unlabeled data.

The most traditional class of SSL methods is based on the manifold regularization [1] which assumes that examples with similar features tend to have similar label attributes. Manifold regularization aims to explore the local geometry of data distribution using a penalty term to steer the regression function along the potential manifold. Representative manifold regularization includes locally linear embedding (LLE) [2], ISOMAP [3], local tangent space alignment (LSTA) [4],

* Corresponding author. Tel.: +86 13560476207.
E-mail address: dtao.scut@gmail.com (D. Tao).

http://dx.doi.org/10.1016/j.sigpro.2014.08.002
0165-1684/

Laplacian regularization (LR) [5], and Hessian regularization (HR) [6]. Particularly, LLE uses linear coefficients to represent the local geometry. ISOMAP is a variant of MDS [7] which preserves global geodesic distances of all pairs of measurements. LSTA represents the local geometry by exploiting the local tangent information. LR develops approximate representation on Laplacian graph. HR obtains the presentation by estimating the Hessian over neighborhood. Comparing with other manifold regularization methods, Hessian has richer nullspace and drivers the learned function varying linearly along the underlying manifold [6,8,9,25]. Thus Hessian regularization is preferable for encoding the local geometry and then boosts the SSL performance.

On the other hand, multimedia is naturally represented by multiview features [21,22,45–48,50] including spatial and time information. Different views character different data properties and features for different views are often complimentary to one another. Thus multiview features can provide more characteristics and significantly improve the learning performance. Popular multiview learning algorithms include co-training [10], multiple kernel learning (MKL) [11–14], and graph ensemble learning [15–20]. Co-training based algorithms learn from two different views to exploit unlabeled data to improve the classification performance. MKL learns a kernel machine from multiple Gram kernel matrices which are built from different view features. Graph ensemble based methods explore the complementary properties of different views by integrating multiple graphs which encode the local geometry of different views.

Human action recognition is one typical and challenging job in computer vision and has many applications including gait recognition [29], group activity analysis [30], abnormal activity detection [31], and sport activity recognition [32]. There has been various research efforts reported based on this scenario [42], which focus on activity feature extraction [33–35], human activity representation [36,37,49] and related classification algorithms [38–41,53]. Recently, several researchers are pursuing issues of multiview characters and manifold information to promote the development of human action recognition [43,44].

Based on the concerns above, in this paper we present multiview Hessian regularized logistic regression (mHLR) for action recognition. Significantly, mHLR seemly integrates multiview features and Hessian regularizations constructed from different views. The merits of mHLR lie in three folds: (1) mHLR combines multiple Hessian regularization to drive the leaned function varying linearly along the underlying manifold and then leverages the exploring of local geometry; (2) mHLR naturally handles multi-view instances with multiple representations to boost action recognition; (3) mHLR employs a smooth loss function and then can be effectively optimized. We conduct action recognition experiments on the unstructured social activity attribute (USAA) dataset [23,24]. The experimental results show the effectiveness of mHLR by comparing with the baseline algorithms.

The rest of this paper is assigned as follows. Section 2 presents the proposed multiview Hessian regularized logistic regression. Section 3 details the optimization algorithm of mHLR. Section 4 discusses the experimental results on USAA dataset, followed by the conclusion in Section 5.

2. Multiview Hessian regularized logistic regression

In multiview Hessian regularized logistic regression, we are given l labeled examples $\mathcal{L}=\{(x_i^1,x_i^2,\ldots,x_i^V,y_i)\}_{i=1}^{l}$ and u unlabeled examples $\mathcal{U}=\{(x_i^1,x_i^2,\ldots,x_i^V)\}_{i=l+1}^{l+u}$, where V is the number of views, $x_i^k\in\mathcal{X}^k$ for $k\in\{1,2,\ldots,V\}$ is the feature vector of the kth view of the ith example, and $y_i\in\{\pm 1\}$ is the label of example x_i. In the rest section of this paper, we use $x_i=\{x_i^1,x_i^2,\ldots,x_i^V\}$ to represent the ith example and x^k to represent the kth view feature. Both labeled and unlabeled examples are drawn from the underlying manifold $\mathcal{M}$. Typically, $l\ll u$ and the goal is to predict the labels of unseen examples.

By incorporating a manifold regularization term to control the complexity of the learning function, the mHLR can be written as the following optimization problem:

$$\min_{f\in\mathcal{H}_K}\frac{1}{l}\sum_{i=1}^{l}\varphi(f,x_i,y_i)+\gamma_K\|f\|_K^2+\gamma_I\|f\|_I^2 \tag{1}$$

where $\varphi(\cdot)$ is the logistic loss function, $\|f\|_K^2$ penalizes the classifier complexity in an appropriate reproducing kernel Hilbert space (RKHS) $\mathcal{H}_K$, $\|f\|_I^2$ controls f along the compact manifold $\mathcal{M}$, and γ_K and γ_I are parameters which balance the loss function and regularization terms $\|f\|_K^2$ and $\|f\|_I^2$ respectively.

As mentioned above, examples are represented by multiview features. Then the proposed mHLR integrates multiple kernel learning and ensemble graph Hessian learning.

Suppose K^k, $k=1,2,\ldots,V$ is a valid (symmetric, positive definite) kernel on the kth view, and then we define the multiview kernel as follows:

$$\boldsymbol{K}=\sum_{k=1}^{V}\theta^kK^k,\quad s.t.\ \sum_{k=1}^{V}\theta^k=1,\quad \theta^k\ge 0,\qquad k=1,2,\ldots,V.$$

Suppose $K^k=\langle\phi_k(\omega^k),\phi_k(\mu^k)\rangle:\mathcal{X}\times\mathcal{X}\to R$, then $\theta^kK^k=\langle\sqrt{\theta^k}\phi_k(\omega^k),\sqrt{\theta^k}\phi_k(\mu^k)\rangle$.

Then we have

$$\begin{aligned}\boldsymbol{K}(\omega,\mu)&=\sum_{k=1}^{V}\theta^kK^k(\omega^k,\mu^k)\\&=\sum_{k=1}^{V}\left\langle\sqrt{\theta^k}\phi_k(\omega^k),\sqrt{\theta^k}\phi_k(\mu^k)\right\rangle\\&=\left\langle\left[\sqrt{\theta^1}\phi_k(\omega^1)\cdots\sqrt{\theta^k}\phi_k(\omega^k)\cdots\sqrt{\theta^V}\phi_k(\omega^V)\right],\right.\\&\qquad\left.\times\left[\sqrt{\theta^1}\phi_k(\mu^1)\cdots\sqrt{\theta^k}\phi_k(\mu^k)\cdots\sqrt{\theta^V}\phi_k(\mu^V)\right]\right\rangle.\end{aligned}$$

Therefore, the multiview kernel $\boldsymbol{K}$ is also a valid (symmetric, positive definite) kernel. And the regularization term $\|f\|_K^2=\mathbf{f}^T\boldsymbol{K}\mathbf{f}=\sum_{k=1}^{V}\theta^k\|f\|_{K(k)}^2$.

Similarly we define multiview Hessian. Suppose $H^j, j=1,2,\ldots,V$ is the Hessian of the jth view, we have

$$\boldsymbol{H}=\sum_{j=1}^{V}\beta^jH^j,\quad s.t.\ \sum_{j=1}^{V}\beta^j=1,\quad \beta^j\ge 0,\qquad j=1,2,\ldots,V.$$