Manifold-Regularized Selectable Factor Extraction for Semi-supervised Image Classification

Xin Shi[1]
shixinnk@gmail.com
Chao Zhang[1]
chzhang@cis.pku.edu.cn
Fangyun Wei[1]
weifangyun@pku.edu.cn
Hongyang Zhang[1]
hy_zh@pku.edu.cn
Yiyuan She[2]
yshe@stat.fsu.edu

[1] Key Laboratory of Machine Perception (MOE), School of EECS, Peking University, PR China
[2] Department of Statistics, Florida State University, USA

Feature selection methods are efficient in modern computer vision applications to reduce the computational cost and the chance of over-fitting. Recently, a novel selectable factor extraction (SFE[3]) framework is proposed to simultaneously perform feature selection and extraction, and is theoretically and practically proved to be effective for high-dimensional data. Although it is advantageous in several aspects, SFE is only designed for either supervised or unsupervised learning, and is not suitable when there are limited labeled samples and a large number of unlabeled samples. To tackle this problem, we propose a novel manifold regularized SFE (MRSFE) framework for semi-supervised image classification.

We use a low rank penalized regression model to explore the label information. A low rank matrix of the regression coefficients, together with the $\ell_{2,1}$ or $\ell_{2,0}$ norm penalty is learned for joint feature selection and extraction. In addition, all the labeled and unlabeled samples are utilized in MRSFE to construct the data adjacency graph to approximate the underlying data manifold, which the data distribution is assumed to be supported on. The graph Laplacian is then incorporated as a regularization term to smooth the coefficients matrix. In this way, the local structures of the whole dataset are preserved, and the data distribution is well exploited.

To derive our model, we begin with the reduced rank regression (RRR) model[1]:

$$\min_B \|Y_L - X_L B\|_F^2, \quad \text{s.t. } \operatorname{rank}(B) \le r. \tag{1}$$

where $B = [b_1, b_2, \cdots, b_d]^T \in \mathbb{R}^{d\times c}$ is the representation matrix, $X_L \in \mathbb{R}^{l\times d}$ is the design matrix of the labeled samples, l is the number of the labeled samples and d is the number of the features, $Y_L \in \mathbb{R}^{l\times c}$ is the response matrix where c is the number of the classes, i.e., $Y_{ij} = 1$ if and only if the ith sample belongs to the jth class.

Although successful in practice, the plain RRR has however certain drawbacks — the model (1) typically involves all input features of X. To this end, we implement feature selection for RRR,

$$\min_B \|Y_L - X_L B\|_F^2 + \alpha\|B\|_{2,1}, \quad \text{s.t. } \operatorname{rank}(B) \le r. \tag{2}$$

where $\|B\|_{2,1} = \sum_{i=1}^d \sqrt{\sum_{j=1}^c B_{ij}^2}$, which promotes row sparsity of B. Note that we select features from X_L according to the non-zero rows of the representation matrix B, i.e., if the i-th row of B is non-zero, we conclude that the ith feature of X_L(i-th column) is significant. We call (2) the selectable factor extraction (SFE) method.

In many applications where we have only a small number of labeled samples, the representation matrix B learned from model (2) is often unreliable. We use the manifold regularization (MR[2]) to help learn the manifold structure of large number of unlabeled samples. Combining model (2) with MR, we have the following optimization problem:

$$\min_B \|Y_L - X_L B\|_F^2 + \alpha\|B\|_{2,1} + \beta\operatorname{tr}(B^T X_{LU}^T L X_{LU} B), \text{s.t. } \operatorname{rank}(B) \le r. \tag{3}$$

where L is the Laplacian matrix, X_{LU} is the data matrix containing both the labeled and unlabeled samples.

To deal with the row sparsity and the rank constraint on B, we write $B = SV^T$, where $S \in \mathbb{R}^{d\times r}$ and $V \in \mathbb{R}^{c\times r}$ is an orthogonal matrix.Thus model (3) is equivalent to

$$\min_{S, V^TV=I_r} \|Y_L - X_L SV^T\|_F^2 + \alpha\|S\|_{2,1} + \beta\operatorname{tr}(VS^T X_{LU}^T L X_{LU} SV^T). \tag{4}$$

The full-rank factorization of B enables us to tackle the sparsity regularization and the low-rank constraint separately. In other words, S selects significant features from X_L, while the orthogonal matrix V determines the subspace after dimension reduction. So we can implement feature selection and extraction simultaneously.

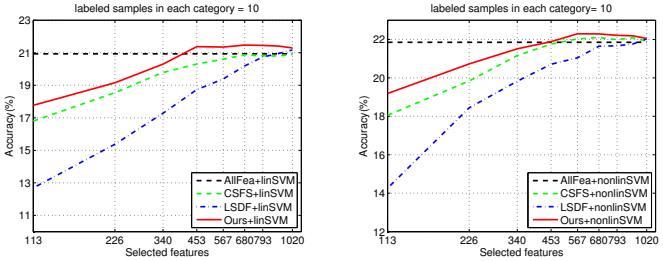

Figure 1: Prediction accuracy vs. the number of selected features (left) using linear SVM (right) using nonlinear SVM

Considering the ℓ_1 penalty cannot handle the collinearity and may lead to inconsistent and biased estimation([4]), similar to $\ell_{2,1}$, we advocate to use non-convex constraint such as $\ell_{2,0}$ instead of widely-used $\ell_{2,1}$ penalty for S in (4)

$$\min_{S, V^TV=I_r} \|Y_L - X_L SV^T\|_F^2 + \beta\operatorname{tr}(VS^T X_{LU}^T L X_{LU} SV^T), \text{s.t. } \|S\|_{2,0} \le q_s \tag{5}$$

where q_s is a parameter to control the number of selected features. Using the constraint form instead of the penalty is intuitive, because we can directly control the number of features we need. Although the $\ell_{2,0}$ penalty is nonconvex and hard to optimization in classical methods, it is doable in our algorithm. Once S is obtained from (4) or (5), we can select the significant features according to top-k index of the row-norms in descending order or nonzero rows of S. We call both of the models (4) and (5) manifold-regularized semi-supervised selectable factor extraction method(MRSFE).

We use alternating optimization method to solve model (4) and (5), an efficient and easy-to-implement algorithm is designed to find the solutions. Our algorithm only consists of SVD decomposition of small-scale W, together with some thresholding operations which are at low cost. We evaluate the effectiveness of our MRSFE by applying it to a challenge web image dataset, NUS-WIDE-OBJECT. Experiments on a this dataset demonstrate the superiority of the proposed method.

[1] Alan Julian Izenman. Reduced-rank regression for the multivariate linear model. *JMA*, 5(2):248–264, 1975.

[2] Yong Luo, Dacheng Tao, Bo Geng, Chao Xu, and Stephen J Maybank. Manifold regularized multitask learning for semi-supervised multilabel image classification. *IEEE TIP*, 22(2):523–536, 2013.

[3] Yiyuan She. Selectable factor extraction in high dimensions. *arXiv preprint arXiv:1403.6212*, 2014.

[4] Peng Zhao and Bin Yu. On model selection consistency of lasso. *JMLR*, 7:2541–2563, 2006.