Joint Feature Selection with Low-rank Dictionary Learning

Homa Foroughi
homa@ualberta.ca

Moein Shakeri
shakeri@ualberta.ca

Nilanjan Ray
nray1@ualberta.ca

Hong Zhang
hzhang@ualberta.ca

Department of Computing Science
University of Alberta
Edmonton, Canada

In many areas, such as computer vision and pattern recognition, data are characterized by high dimensional feature vectors. If these vectors are processed directly, it usually leads to difficult pattern recognition task. However, in practice, only a small subset of features is really important and discriminative. Feature selection (FS) is one of the well known dimensionality reduction methods that efficiently describes the input data by removing irrelevant variables and reduces the effects of noise to provide good prediction results.

Recently, Yan *et al.* [2] introduced the sparse representation-based classification (SRC) [1] measurement criterion into FS and designed a joint sparse discriminative FS method. Based on the assumption of SRC, their method selects a subset of features which minimize the within-class reconstruction residual and simultaneously maximize the between-class reconstruction residual in the subset of selected features. Although they can achieve promising results compared to other FS methods, it is well-known that SRC suffers from major drawbacks such as high computational complexity and more importantly, low discriminativity of sparse coefficients and naive dictionary. Consequently, the reconstruction scatter matrices would not preserve the reconstructive relationship of data well, which means the selected features are not discriminant enough.

To overcome the drawbacks associated with the SRC algorithm, in this paper we propose a new FS method by learning a smaller-sized dictionary while maintaining the sparse reconstruction relationship among samples. In order to generate discriminative dictionary and sparse coefficients, we utilize the power of a supervised dictionary learning (DL) method called Fisher discrimination dictionary leaning (FDDL) [3]. The discrimination capability of FDDL originates from two facts. First, each sub-dictionary is trained to have good representation power to the samples from the corresponding class, but have poor representation power to the samples from other classes. Second, the sparse coefficients are made discriminative through minimizing the within-class scatter and maximizing the between-class scatter of them. Both of these properties make FDDL a good choice for finding the discriminative sparse coefficients of training samples. Nevertheless, all the DL algorithms, including FDDL, work well when the input images are clean or corrupted by small noise and their performance deteriorates when the training data is contaminated because of occlusion, disguise or lighting variations [4]. Low-rank (LR) matrix recovery, which determines a LR data matrix from corrupted data, has been successfully applied to different tasks including image classification. To improve the performance of FDDL with noises, we integrate LR matrix approximation into sparse representation for DL. Specifically, we formulate FS problem under LR dictionary learning with Fisher discrimination regularization and propose a Joint Feature Selection method using Low-rank Dictionary Learning (JFS-LDL) as follows:

(a) **Low-rank Approximation**: Given a set of training data vectors $X = [X_1, X_2, \ldots, X_K]$, the samples from i^{th} class, X_i, are linearly correlated in many situations. LR matrix recovery seeks to decompose a data matrix X_i into $L_i + E_i$ by minimizing the rank of matrix L_i, while reducing $\|E_i\|_0$, the associated sparse noise [4]. Clearly, LR reduces the diversity across items within each class and consequently dissimilarity between different classes is increased, which means sub-dictionaries are more discriminant toward each other.

(b) **Low-rank Dictionary Learning using Fisher Discrimination**: To improve the performance of FDDL with noises and contamination, we use a more robust representation of X_i, which is basically its LR representation, L_i in the FDDL objective function. Furthermore, when the standard LR matrix recovery is combined with

Fisher discrimination, the images tend to be more similar to each other for the same class, which means more compactness exists within the same class and dissimilarity between different classes. Therefore, the learned sub-dictionaries would have better discrimination and reconstruction capabilities compared to FDDL model. Accordingly, the quality of structured dictionary will influence the discriminativeness of the sparse coefficients.

(c) **Joint Feature Selection**: We aim to select a subset of features that preserve the sparse reconstructive relationship of the training samples. This is achieved by minimizing the within-class reconstruction residual error and simultaneously maximizing the between-class reconstruction residual in the subset of selected features. We exploit the sparse coefficients obtained by objective function of LR dictionary learning using Fisher discrimination to obtain the within-class and between-class reconstructive scatter matrices. Simultaneously, the $l_{2,1}$-norm minimization on projection matrix is applied to jointly select the most relevant and discriminative features. As a result, the projected samples are more discriminative and simultaneously retain the important properties for classification, e.g., intraclass compactness and interclass separability, as well as the reconstructive relationship.

We conduct extensive experiments on benchmark datasets including handwritten digits, face and sport action to verify the effectiveness of the proposed JFS-LDL in comparison with other FS methods and validate its capability for image classification task. Experiments show that JFS-LDL consistently outperforms all the other evaluated FS methods, especially in lower dimensions in image classification task. JFS-LDL can maintain a relatively stable performance under different dimensions, and as the number of selected features decreases, its advantage becomes more obvious. The combination of LR approximation and Fisher discrimination, leads in more compactness within the same class and dissimilarity between different classes. Consequently, a simple and fast classifier like KNN or SVM would perform well for classification. In contrasts to most of DL methods which use l_1-optimization to find the representation of test images and use the reconstruction error for classification, our classification schema is very efficient and fast. Furthermore, we observe that JFS-LDL with a selected subset of features can achieve superior or competitive recognition rate compared to the recently proposed DL methods, with much higher feature dimensions. This implies the effectiveness of our method in capturing the discriminative information for classification. The experimental results together with the theoretical analysis validate the effectiveness of our method for feature selection, and its efficacy for image classification.

[1] John Wright, Allen Y Yang, Arvind Ganesh, Shankar S Sastry, and Yi Ma. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2): 210–227, 2009.

[2] Hui Yan and Jian Yang. Sparse discriminative feature selection. *Pattern Recognition*, 2014.

[3] Meng Yang, Lei Zhang, Xiangchu Feng, and David Zhang. Sparse representation based fisher discrimination dictionary learning for image classification. *International Journal of Computer Vision*, 109(3): 209–232, 2014.

[4] Yangmuzi Zhang, Zhuolin Jiang, and Larry S Davis. Learning structured low-rank representations for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 676–683. IEEE, 2013.