Exploiting Low-rank Structure for Discriminative Sub-categorization

Zheng Xu[1]
xuzhustc@gmail.com
Xue Li[1]
echolixue@gmail.com
Kuiyuan Yang[2]
kuyang@microsoft.com
Tom Goldstein[1]
tomg@cs.umd.edu

[1] Department of Computer Science,
University of Maryland,
College Park, USA

[2] Microsoft Research,
Beijing, China

In visual recognition, sub-categorization, which divides a category into some sub-categories, has been proposed to deal with large intra-class variance in the real world. Recent discriminant sub-categorization approaches utilize samples that do not belong to the category under consideration as negative data for supervision, and cluster positive samples of the category into sub-categories, then simultaneously train the corresponding classifier for each sub-category [2, 4]. In the jointly clustering and classification framework of previous methods, the classifier for each sub-category is trained by using samples hard-assigned to the sub-category. However, some samples would contribute to the training of several sub-categories since the intra-variance of a category is caused by complex factors. Moreover, sub-categories are closely related since they are discovered from the same category, and the common information among these sub-categories is beneficial for classifier training.

We propose a new approach for discriminative sub-categorization, which adopts the exemplar based method to address the intra-variance in category, and exploits the low rank structure to preserve common information while discovering sub-categories. Our approach builds up the exemplar-LDAs [3], which generates a set of exemplar classifiers with each classifier trained by a single positive sample and all the negative samples. The extreme case of sub-category is to have only one positive sample, which is a compact set for training and modeling. We adopt exemplar classifiers to represent the compact sub-categories and preserve intra-variance in a category. In order to share common information among exemplar classifiers while preserving diversity, we jointly train the exemplar-LDAs for all the positive samples and introduce the trace-norm regularizer on the matrix of weights, as we assume the weights lie on a union of subspaces such that the matrix of weights is low-rank.

We formulate the proposed low-rank least squares exemplar-LDAs (LRLSE-LDAs) as follows. Let $\mathbf{X}_1 = [\mathbf{x}_1^+, \dots, \mathbf{x}_n^+]$ and $\mathbf{X}_2 = [\mathbf{x}_1^-, \dots, \mathbf{x}_m^-]$ denote the centered data matrix[1] for positive samples and negative samples, $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_n]$ denote the weight matrix where each w_i is the weight vector of exemplar-LDA for a positive sample. The objective function for training the exemplar-LDAs of positive samples together is

$$J_{LSE-LDAs}(\mathbf{W}) = \frac{\delta}{2}\|\mathbf{W}\|_F^2 + \frac{1}{2}\|\mathbf{X}_2'\mathbf{W}\|_F^2 - trace(\mathbf{X}_1'\mathbf{W}) \quad (1)$$

where $\|\cdot\|_F$ is the Frobenius norm of a matrix, $trace()$ represents the trace of a matrix. We minimize the least squares form in Eq. 1 instead of maximizing the Fisher criterion so that the objective function is convex, inspired by [6]. Eq. 1 has closed-form solution as

$$\mathbf{W} = (\mathbf{X}_2\mathbf{X}_2' + \delta\mathbf{I})^{-1}\mathbf{X}_1 \quad (2)$$

where $\mathbf{I}$ is the identity matrix. To discover the structure of sub-categories, we jointly learn the weight for positive samples/exemplars of the category and regularize the weight matrix with a low-rank constraint. Finally, we arrive at the objective function of LRLSE-LDAs,

$$J_{LRLSE-LDAs}(\mathbf{W}) = \xi\|\mathbf{W}\|_* + J_{LSE-LDAs}(\mathbf{W}) \quad (3)$$

$\|\cdot\|_*$ is the trace norm used to regularize the weight matrix, which is a convex approximation of the rank of a matrix

To solve the convex formulation in Eq. 3, we propose an efficient algorithm based on the scaled form of alternating direction method of multipliers (scaled ADMM) [1]. We reformulate minimizing $J_{LRLSE-LDAs}(\mathbf{W})$ in Eq. 3 as an equality-constrained convex optimization problem by introducing an intermediate variable $\mathbf{F}$,

$$\min_{\mathbf{W},\mathbf{F}} J_{LSE-LDAs}(\mathbf{W}) + \xi\|\mathbf{F}\|_* \quad \text{s.t.} \quad \mathbf{W} = \mathbf{F} \quad (4)$$

The augmented Lagrangian for the formulation in Eq. 4 can be written as:

$$L(\mathbf{W},\mathbf{F},\mathbf{\Lambda}) = J_{LSE-LDAs}(\mathbf{W}) + \xi\|\mathbf{F}\|_* + \tfrac{\tau}{2}(\|\mathbf{W}-\mathbf{F}+\mathbf{\Lambda}\|_F^2 - \|\mathbf{\Lambda}\|_F^2) \quad (5)$$

where $\mathbf{\Lambda}$ is the scaled dual parameter matrix, and τ is the penalty parameter. We iteratively update variables $\mathbf{W},\mathbf{F},\mathbf{\Lambda}$ as in scaled ADMM, where $\mathbf{W},\mathbf{F}$ are updated by solving two subproblems both with closed-form solutions, and $\mathbf{\Lambda}$ is updated by dual ascent. The two subproblems are

$$\mathbf{W} = \arg\min_{\mathbf{W}} J_{LSE-LDAs}(\mathbf{W}) + \frac{\tau}{2}\|\mathbf{W}-\mathbf{F}+\mathbf{\Lambda}\|_F^2 \quad (6)$$

$$\mathbf{F} = \arg\min_{\mathbf{F}} \xi\|\mathbf{F}\|_* + \frac{\tau}{2}\|\mathbf{W}-\mathbf{F}+\mathbf{\Lambda}\|_F^2 \quad (7)$$

where Eq. 6 has a closed-form solution benefits from the least squares form and Eq. 7 can be solved by singular value thresholding method.

After training the weights of LRLSE-LDAs, we utilize those exemplar classifiers to perform sub-category discovery and visual recognition. For sub-category discovery, we adopt spectral clustering with affinity matrix defined by the prediction scores on positive samples. For visual recognition, we adopt the cross domain recognition approach in [5] by fusing the top-K prediction scores from trained exemplar classifiers.

We conduct comprehensive experiments on various datasets to validate the effectiveness and efficiency of our approach in sub-category discovery and visual recognition. We follow the experimental setting in [4] to evaluate the performance of sub-category discovery. We conduct experiments on ten public datasets from the UCI repository and MNIST, which cover a large variant types of data. LRLSE-LDAs based clustering achieves promising results measured by purity on those datasets. We follow the experimental setting in [5] to evaluate the performance of visual recognition. We use the Office-Caltech dataset for object recognition and the IXMAS dataset for action recognition. LRLSE-LDAs based classification achieves order-of-magnitude speedup with matching performance comparing with state-of-the art in [5].

[1] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[2] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1627–1645, 2010.

[3] Bharath Hariharan, Jitendra Malik, and Deva Ramanan. Discriminative decorrelation for clustering and classification. In *ECCV*. 2012.

[4] Minh Hoai and Andrew Zisserman. Discriminative sub-categorization. In *CVPR*, 2013.

[5] Zheng Xu, Wen Li, Li Niu, and Dong Xu. Exploiting low-rank structure from latent domains for domain generalization. In *ECCV*, 2014.

[6] Jieping Ye. Least squares linear discriminant analysis. In *ICML*, 2007.

[1] Data matrix is centered by subtracting the mean of training samples from each sample. We use mean of negative samples to approximate the mean of all negative sample and a positive sample for each exemplar classifier.