SMLFDL

Note: These are reports and reviews which must not be used as refernece for other works

1st A.H.Ebrahimi
dep. Statistics, Mathematics and Computer Science
University of Tehran
Birjand, Iran
ah.ebrahimi@ut.ac.ir

2nd Dr. H.Sajedi
dept. Statistics, Mathematics and Computer Science
University of Tehran
Tehran, Iran
hediehsajedi@gmail.com

***Abstract*—Classifying images is a complex problem in the field of computer vision, it gets even harder for multi-class classification. Recently, There were an increasing interest in learning model due to its promising results in various image classification tasks, but downside for their need of large amount of training data, they can not achieve a high accuracy of classification with a small-size data-set. In contrast, dictionary learning (DL) methods can achieve a perfect performance on a image classification task and hence still get a lot of attention. Among these DL methods, DL based feature learning methods are the mainstream for image classification in recent years, however, most of these methods have trained a classifier independently from dictionary learning. Therefore, the features extracted by the learned dictionary may not be very proper to perform classification for the classifier [1]. On the other hand SVMs multi-class loss feedback based discriminative dictionary learning (SMLFDL) try to learn a discriminative dictionary and SVM coeffinencents at the same time. This method which has been inspired by the feedback mechanism in cybernetics makes the dictionary features and SVMs better matched with each other. Because of this unified learning framework better classification performance can be achieved compare to state-of-the-art dictionary learning methods.**

***Index Terms*—DL, SVM, cybernetics**

I. Introduction

Image classification is one of the most challenging problems in machine learning, especially with a large number of object categories. It is because that there is a great deal of variability in the involved images, such as illumination, viewpoint variations, occlusion and corruption [2].

There has been lots of researches about finding appropriate features of image, for instance, Gabor, LBP, SIFT, Daisy and HOG feature descriptors. Although, these low-level features can perform well in some image classification task, they can not easily be adopted to new variability of new data that is sampled under other new conditions, that is characteristic of their manually design process Therefore, learning features from the original data is regarded as a promising approach to surmount the shortcoming of these hand-crafted features [2].

II. Dictionary Learning

The key step of classifying images is obtaining feature representations encoding relevant label information. In the last decade, one of the most popular representation learning methods is Dictionary learning also called sparse representation or sparse coding. Dictionary learning can learn a set of atoms so that a given signal (or possibly an image) can be well approximated by a sparse linear combination of these learned atoms. As mentioned Dictionary Learning represents big amount of data encapsulated in a signal or an image with a linear combination of a few atoms (elements) from a dictionary matrix which contains several prototype signal-atoms. The coefficient vector specifies how to linearly combine and remake original signal from dictionary.

Some of fundamental advantages of sparse representation are simplifying the complex signal and making the classification procedure more tractable. Signals such as audios, videos and images admit naturally sparse representations, while the key idea of sparse representation is mapping the original chaotic signals to their corresponding concise representations with a regularized or uniform style [3].

III. Classification scheme of the SMLFDL

The dictionary D and the classifiers ¡w, b¿ are both obtained under the proposed SMLFDL framework. Given a test sample x, it can be coded with the following representation model:

$$\hat{\alpha} = argmin_{\alpha}\{||x - D||_2^2 + \lambda_1||\alpha||_2^2\} \tag{1}$$

Obviously, the closed solution of the equation 1 can be formulated by the following:

$$\hat{\alpha} = (D^T D + \lambda_1)^{-1} D^T x \tag{2}$$

Through the solved coefficient vector $\hat{\alpha}$, the C linear classifiers $< w_c, b_c >$ where $c \in \{1, 2, \dots, C\}$ can be used to make the decision of the sample x's label by:

$$y = argmax_{c\in\{1,2,\dots,C\}} w_c^T \hat{\alpha} + b_c \tag{3}$$

A. SMLFDL Algorithm

There are three main procedures involved in SMLFDL algorithm: solving/updating the coefficients, training/updating SVMs and updating the dictionary [1]. Training data are passed trough a tunnel which starts with Dictionary model, in this stage data are used to make a dictionary atoms D and a coefficients Z. The Dictionary D will be used latter on classification by the equation 2. Coefficients are passed to C linear SVM where C is number of unique classes which has been used. Coefficients are classified by one-vs-other strategy in SVMs training algorithm. For SVMs intercepts and coefficients update one can use L-BFGS minimum optimization algorithm which results faster convergence compared with old gradient descent. SVMs $< w_i, b_i >$ are going to be used in final prediction by the equation 3. SMLFDL is an iterative algorithm and for updating SVM intercepts and coefficients beside learned dictionary, B.Q Yang et al. proposed a multi loss function based on the Lagrange dual algorithm which construct a Fisher discrimination term on the reconstruction penalty term. This multi-objective-loss function achieved a better classification result [4]. SMLFDL integrates dictionary learning and SVMs training together. Therefore, the dictionary D and SVMs can be obtained together under a unified learning framework so that the coefficients as features extracted by D and SVMs will be better adapted with each other. In addition, the employment of the Fisher criterion will promote intra-class coefficient compactness and inter-class coefficient separability in the SMLFDL model. Consequently, SMLFDL will achieve a better pattern classification as well as be robust to the intra-class variability of data-sets to some extent [1].

1) Update $< w, b >$ of SVMs: When providing the coefficients Z, the update of $< w, b >$ can be formulated to train C linear SVMs by one-against-all. one can use LBFGS method to update SVMs intercepts and coefficients in next iterations to improve speed for update. The L-BFGS method solves the unconstrained minimization problem

$$minimize F(x), x = (x1, x2, ..., xN)$$

Only if the objective function F(x) and its gradient G(x) are computable. The well-known Newton's method requires computation of the inverse of the hessian matrix of the objective function. However, the computational cost for the inverse hessian matrix is expensive especially when the objective function takes a large number of variables. The L-BFGS method iteratively finds a minimizer by approximating the inverse hessian matrix by information from last m iterations. This innovation saves the memory storage and computational time drastically for large-scaled problems [5].

2) Update dictionary D: The dictionary D of SMLFDL and the coefficients Z of training samples over D are updated via

$$\begin{aligned}\langle D, Z\rangle = \arg\min_{D,Z} \|X - DZ\|_F^2 + \lambda_1 \|Z\|_p^p + \lambda(tr(S_W(Z)) - tr(S_B(Z))) \\ + \gamma \sum_{i=1}^{n} \sum_{c=1, c\neq y_i}^{C} [\max(0, w_c^T z_i + b_c - w_{y_i}^T z_i - b_{y_i} + \epsilon)]^2 \\ \text{s.t.} \quad \|d_k\|^2 = 1, \forall k \in \{1, 2, \ldots, K\}\end{aligned}$$

This optimization problem could be solved by alternative and iterative processes: updating D while fixing Z; updating Z while fixing D [1].

3) Update coefficients Z: Keeping D fixed and omitting irrelevant terms, the objective function could be simplified.l_2-norm regularizer was adopted on the coefficients, preventing not being convex and unstablility of objective. Fisher discriminant instaed of norm 2, to make the objective function smoother and convex. The coefficients $Z = [z1, z2, \ldots, zn]$ could be optimized column by column.

- Calculate probable classes for every i except itself, by SVMs coefficient and intercepts.
- Update Z columns conditionally base on existence of such a probable class.

Algorithm 1: SVM Multi-class Loss Feedback Based Dictionary Learning Algorithm.

input : X, D$_{init}$, $Z_{init}, \lambda, \lambda_1 and \gamma$.
output: D, w and b.
while *not converge or not reaching the maximal iteration steps* **do**
  train C linear SVM by ova strategy;
  update Z column by column based on existence of probable class;
  update D by dictionary learning algorithm;
end

The atoms of each sub-dictionary D_c are initialized as the eigenvectors of training samples X_c through principal component analysis (PCA) algorithm, and these D_cs are then concatenated to form the initialized D. Then each column of Z is initialized as an appropriate column vector of a transformation on $D^T x_i$.

consider

$$D^\dagger = D^T D + \lambda_1 I \quad (4)$$

Z columns are updated with the followings schemed formula:

$$(D^\dagger + \theta)^{-1} * (D^T x + \zeta) \quad (5)$$

θ, and ζ are numbers gain from different strategies based on the most probable class except y_i class label. There will be nearly the similar scheme for predictions.

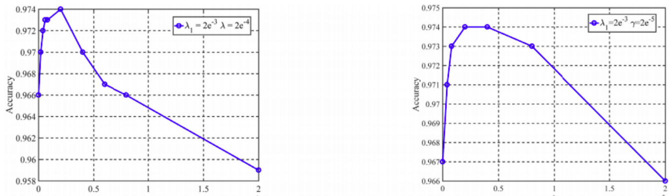
$$(D^\dagger)^{-1} * (D^T x) \quad (6)$$

The only difference of this formula compare to Z columns is the removal of θ, and ζ. Because SVM tries to maximum the marginal distance it would be a great match for our classification task, by classifying coefficients in the most generalized way.

B. Parameter selection

The size of each D_c, denoted by K_c, $c = \{1, 2, \ldots, C\}$, have great influence on the classification accuracy of SMLFDL. Each K_c is usually set to be equal in SMLFDL. In order to assess its influence in SMLFDL [1].

There are λ_1, λ and γ in the DL model of SMLFDL. λ_1, λ and γ are related to the regularization term $||Z||_p^p$, $L(w, b, z_i)$ and $f(Z)$, respectively. $L(w, b, z_i)$ is the multi-class loss term that is looped into the dictionary learning model, and f(Z) is the discriminative coefficient term that is introduced as the Fisher criterion on the representation coefficients.



(a) Accuracy of SMLFDL vs. different γ values on the Extended Yale B database.

(b) Accuracy of SMLFDL vs. different λ values on the Extended Yale B database.

(c) Accuracy of SMLFDL vs. different γ values on Test 2 of the Multi-PIE database

(d) Accuracy of SMLFDL vs. different λ values on Test 2 of the Multi-PIE database

Fig. 1. Accuracy of SMLFDL vs. different and on the Extended [1]

C. Scene classification

In the scene classification task, the proposed SMLFDL is verified on the Scene-15 database. The Scene-15 database consists of 4485 scene images that cover suburb, bedroom, kitchen, industrial, living, room, forest, inside city, coast, office, highway, tall building, mountain, street, open country, and store scene categories. Each category contains the number of images (250 × 300 pixels) varies from 210 to 410 [8].

For the 15 scene category dataset, we used the computed spatial pyramid feature using a four-level spatial pyramid and a SIFT-descriptor codebook with a size of 200. The final spatial pyramid features are reduced to 3,000 by PCA [7]. Then 100 samples per category are randomly selected to train SMLFDL, and the rest are used to test SMLFDL. The size of the dictionary is set to be 750 (50 atoms for per category). λ_1, λ and γ are set to be 2e1, 2e3 and 2e4 respectively in this experiment.

Fig. 2. Some samples from the Fifteen Scene Categories database.

IV. Conclusion and Future Work

we reported the mechanism and algorithm of support vector machines (SVMs) multi-class loss feedback based discriminative dictionary learning (SMLFDL) method for image classification. It seems that outputs of proposed method results in better exactness of the designed multi-class loss and better performance for image classification can [**?**]. It also needs less iterations (on average 20 iteration will be enough) for convergence compared with other existing DL methods. The objective function for this framework used fisher discriminant norm making objective smooth and convex which was not duo its characteristic. SMLFDL simillar to other DL methods need much less training data in contrast with Deep learning methods, also features extracted by these methods are more general and don't need to be manually fitted to data-set.

Acknowledgment

Codes and reports for this SMLFDL was done beside Machine Learning course at The School of Mathematics, Statistics and Computer Science, University of Tehran.

V. Appendix

1) The Spatial pyramid: The technique works by partitioning the image into increasingly fine sub-regions and computing histograms of local features found inside each sub-region. The resulting "spatial pyramid" is a simple and computationally efficient extension of an order less bag-of-features image representation, and it shows significantly improved performance on challenging scene categorization tasks. The spatial pyramid framework also offers insights into the success of several recently proposed image descriptions, including Torralba's "gist" and Lowe's SIFT descriptors and dasiy descriptor.

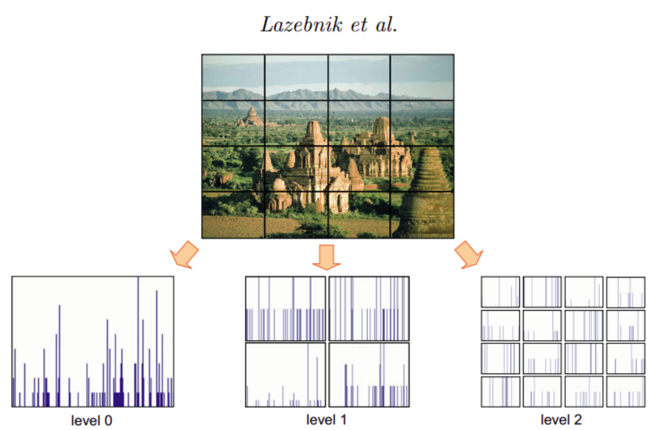


Fig. 3. A schematic illustration of the spatial pyramid representation. A spatial pyramid is a collection of order less feature histograms computed over cells defined by a multi-level recursive image decomposition. At level 0, the decomposition consists of just a single cell, and the representation is equivalent to a standard bag of features. At level 1, the image is subdivided into four quadrants, yielding four feature histograms, and so on. Spatial pyramids can be matched using the pyramid kernel, which weights features at higher levels more highly, reflecting the fact that higher levels localize the features more precisely [6]

2) Dense DAISY feature description: It is possible to estimate depth from two wide baseline images using a dense descriptor. Dense local descriptor (DAISY), is very fast and efficient to compute. It depends on histograms of gradients like SIFT and GLOH but uses a Gaussian weighting and circularly symmetrical kernel. This gives us our speed and efficiency for dense computations. We compute 200-length descriptors for every pixel in an 800×600 image in less than 5 seconds [9].

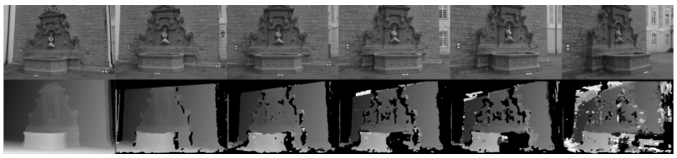


Fig. 4. Stereo reconstructions obtained using the DAISY descriptor.

The first row are the input images where the first image was used and one of the other images as input and the second row shows the respective reconstructions.

REFERENCES

[1] Yang, B. et al. "SVMs multi-class loss feedback based discriminative dictionary learning for image classification." Pattern Recognition (2020): 107690.

[2] T.H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, Y. Ma, PCANet: a simple deep learning baseline for image classification? IEEE Trans. Image Process. 24 (12) (2015) 5017–5032.

[3] Tang, Hao et al. "When Dictionary Learning Meets Deep Learning: Deep Dictionary Learning and Coding Network for Image Recognition with Limited Data." IEEE transactions on neural networks and learning systems PP (2020): n. pag.

[4] Yang, BQ., Gu, CC., Wu, KJ. et al. Simultaneous dimensionality reduction and dictionary learning for sparse representation based classification. Multimed Tools Appl 76, 8969–8990 (2017). https://doi.org/10.1007/s11042-016-3492-1

[5] libLBFGS: a library of Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS), http://www.chokkan.org/software/liblbfgs/index.html

[6] Lazebnik S, Schmid C, Ponce J (2009) Spatial pyramid matching. Object Categorization: Computer and Human Vision Perspectives.

[7] Z. Jiang, Z. Lin, L.S. Davis, Label consistent K-SVD: learning a discriminative dictionary for recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (11) (2013) 2651–2664.

[8] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2006, pp. 2169–2178.

[9] E. Tola, V. Lepetit and P. Fua, "DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 5, pp. 815-830, May 2010, doi: 10.1109/TPAMI.2009.77.