BOOSTED MULTIFOLD SPARSE REPRESENTATION WITH APPLICATION TO ILD CLASSIFICATION

Yang Song[1], *Weidong Cai*[1], *Heng Huang*[2], *Yun Zhou*[3], *Yue Wang*[4], *David Dagan Feng*[1,5,6]

[1]BMIT Research Group, School of IT, University of Sydney, Australia
[2]Department of Computer Science and Engineering, University of Texas, Arlington, USA
[3]Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, USA
[4]Bradley Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, USA
[5]CMSP, Department of EIE, Hong Kong Polytechnic University, Hong Kong
[6]Med-X Research Institute, Shanghai Jiaotong University, China

ABSTRACT

Classification performance with sparse representation is largely affected by the low discriminative power of image features. In this study, we propose a new sparse representation model, namely the Boosted Multifold Sparse Representation (BMSR), to improve the classification performance. By dividing the training set into multiple subsets, sparse representation using one subset is used as a weak classifier. A threefold boosting approach is then designed to combine the multiple weak classifiers to create the final class label. We applied the BMSR method to classify image patches of different interstitial lung disease (ILD) patterns using a publicly available dataset. Promising performance improvement over non-boosted sparse representation is shown.

Index Terms— Sparse representation, boosting, classification

1. INTRODUCTION

Sparse representation (SR) has recently been widely applied to solve various classification problems in medical imaging [1–7]. It works by finding the class of reference dictionary producing the minimum reconstruction error for the test image. Classification performance using SR is typically affected by the discriminative power of the image feature. Misclassification arises when there are large within-class variation and between-class overlapping in the feature space.

To improve the classification performance, the basic SR formulation is often integrated with dictionary learning [1–3, 5], to transform the feature space into a more discriminative coding space. Another approach is to modify the reference dictionary according to the test image [4, 6], to achieve a better reconstruction with the correct reference dictionary. Spatial constraints have also been incorporated into SR [6, 7], to obtain spatially consistent labeling among neighboring regions in an image.

In this work, we propose a Boosted Multifold Sparse Representation (BMSR) method, as an alternative mechanism to improve the classification performance using SR. There are three motivations of our design. First, considering the within-class feature variation and between-class feature overlapping, the selection of training image for creating the reference dictionary is an important factor affecting the classification performance using SR. Selecting a subset of training images with more discriminative features would help to reduce the feature variation/overlapping. Second, while there can be various ways in selecting this subset, we choose to simply divide the entire training set randomly into multiple subsets, and perform the classification in a boosting manner with SR on each subset as a weak classifier. Boosting, such as AdaBoost [8], has been widely popular; but to the best of our knowledge, has yet been integrated with SR as weak classifiers. With boosting, the way of combining weak classifiers is determined based on prior learning, and the final classification is expected to be more accurate than the individual weak classifiers. Third, it is reasonable to expect that different subsets would exhibit varying classification accuracies for different classes. We thus design a multifold approach that fold 1 is to use the same set of weak classifiers for all classes, fold 2 is to apply different selections of weak classifiers for different classes, and fold 3 is to combine the results of the first two folds in a boosting manner as well. Note that the proposed BMSR method can be integrated with any SR-based classifiers as weak classifiers, such as those presented in [1–7]; and with the boosted multifold design, the BMSR method is expected to provide further improved classification.

We applied the proposed BMSR method to classify

This work was supported in part by ARC grants.

ILD tissue patterns in high-resolution computed tomography (HRCT) images. ILD can cause progressive scarring of lung tissues and cause breathing difficulties, and differentiating the various types of tissue patterns is important for treatment. We used a publicly available dataset with annotated region-of-interest (ROI) [9], and the image patches are classified into one of the five categories: normal, emphysema, ground glass, fibrosis, and micronodules.

2. METHOD

2.1. Sparse Representation for Classification

Let's define an image as P, the aim is to classify it as one of the N classes: $L(P) = n \in \{1, ..., N\}$, where $L(P)$ denotes the class label of P. We give a brief description of the standard SR-based classification in this section.

Assume that there are Q_n training images of class n. A reference dictionary $D_n \in \mathbb{R}^{Z \times Q_n}$ is constructed by concatenating the Z-dimensional feature vectors of all Q_n images. The training set D thus contains N reference dictionaries $\{D_n\}$. The label $L(P)$ is then computed by:

$$\begin{aligned} x_n &= \operatorname{argmin}_{x_n} \|f - D_n x_n\|_2^2 \ \ s.t. \ \ \|x_n\|_0 \leq C \\ f'_n &= D_n x_n \\ L(P) &= \operatorname{argmin}_n \|f - f'_n\|_2 \end{aligned} \tag{1}$$

where f is the feature vector of P, and x_n is the sparse coefficient vector corresponding to D_n. We further compute the probability of P being a certain class as:

$$Pr(P, D, n) = \exp\left(-\frac{2\|f - f'_n\|_2}{\sum_{n'=1}^{N} \|f - f'_{n'}\|_2}\right) \tag{2}$$

2.2. Boosted Multifold Sparse Representation

Assume image P is of class 1. Intuitively, it would be classified accurately if the training images in D_1 are similar to P while those in the other dictionaries are dissimilar. However, this property is often not met, since feature vectors could exhibit high variations within the same class and low distinctions between different classes.

We design a BMSR method, in which the training set is divided randomly into a number of subsets and a threefold boosted classification is performed with each subset used as a weak classifier. The details are described as follows.

First, given the training set $\{D_n : n = 1, ..., N\}$, we divide each reference dictionary randomly into M subsets. A dictionary subset is denoted as $D^m = \{D_n^m : n = 1, ..., N\}$. Consider a SR-based classifier on a subset D^m as a weak classifier, and there are a total of M weak classifiers. In a boosting construct, the final class label $L_1(P)$ is obtained as a weighted combination of classification results from multiple iterations:

$$\begin{aligned} &Pr_1(P, D, n) = \textstyle\sum_{t=1}^{T_1} W_1(t) Pr(P, D^{H_1(t)}, n) \\ &L_1(P) = \operatorname{argmax}_n Pr_1(P, D, n) \end{aligned} \tag{3}$$

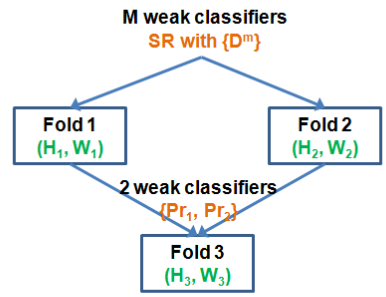

Fig. 1. Method illustration.

where T_1 is the number of iterations, $H_1 \in \{1, ..., m\}^{T_1 \times 1}$ is the selection of weak classifiers, $W_1 \in \mathbb{R}^{T_1 \times 1}$ is the corresponding weight vector, and t indexes the element in H_1 and W_1. A training procedure TrainBoost, similar to AdaBoost, is designed to obtain H_1 and W_1, as listed in Algorithm 1. Here y_i is the ground truth label of training image R_i, $U_x(R_i)$ represents the classification result of R_i using the xth weak classifier and $X = M$. The training outputs h and w are assigned as H_1 and W_1, respectively.

Algorithm 1: Function TrainBoost

Data: training images $\{(R_i, y_i) : i = 1, ..., I\}$, set of weak classifiers $\{U_x : x = 1, ..., X\}$, number of iterations T.
Result: the selected weak classifiers $h \in \{1, ..., X\}^{T \times 1}$, weight vector $w \in \mathbb{R}^{T \times 1}$.

for $i = 1, ..., I$ **do**
  Initialize $K_1(i) = 1/I$;
end
for $t = 1, ..., T$ **do**
  $h_t = \operatorname{argmin}_x \epsilon_{t,x}$;
  $\epsilon_{t,x} = \sum_i^I K_t(i) \mathbf{I}(y_i \neq U_x(R_i))$;
  $w_t = 0.5 \log\{(1 - \epsilon_{t,h_t})/\epsilon_{t,h_t}\}$;
  for $i = 1, ..., I$ **do**
    $K_{t+1}(i) = \dfrac{K_t(i) \exp(w_t(2\mathbf{I}(y_i \neq U_{h_t}(R_i)) - 1))}{\sum_i^I K_t(i) \exp(w_t \mathbf{I}(y_i \neq U_{h_t}(R_i)))}$;
  end
end

Second, we expect that the dictionary subsets would contribute differently for different image classes. In other words, a certain D^m might produce good classification for images of a certain class but many misclassifications for another class. Hence, the second fold is to have a matrix of weak classifiers $H_2 \in \{1, ..., m\}^{T_2 \times N}$ and a weight matrix $W_2 \in \mathbb{R}^{T_2 \times N}$, with each element corresponding to the tth iteration and nth class. Training is conducted by running TrainBoost N times with the M SR-based weak classifiers. At the nth time, only the training images of class n are used, and the outputs α and h are stored as column vectors in W_2 and H_2. At testing time, since the class label of image P is unknown, we would not know in advance which set (column) of H_2 and W_2 to use. Therefore, we iterate through all N possible sets, and label P to the class producing the largest margin δ:

$$\begin{aligned}
&Pr(P, D, n'|n) = \textstyle\sum_{t=1}^{T_2} W_2(t, n) Pr(P, D^{H_2(t,n)}, n') \\
&\delta(P, n) = Pr(P, D, n|n) - \min_{n'} Pr(P, D, n'|n) \\
&L_2(P) = \operatorname{argmax}_n \delta(P, n) \\
&Pr_2(P, D, n) = Pr(P, D, n|L_2(P))
\end{aligned} \tag{4}$$

where $|n$ represents that P is assumed to belong to class n.

Third, the classification procedures in the first two folds are used as two weak classifiers, and the final class label $L(P)$ is obtained by combining the two weak classifiers in a boosting manner:

$$\begin{aligned}
&Pr_3(P, D, n) = \textstyle\sum_{t=1}^{T_3} W_3(t) Pr_{H_3(t)}(P, D, n) \\
&L(P) = \operatorname{argmax}_n Pr_3(P, D, n)
\end{aligned} \tag{5}$$

where $H_3 \in \{1, 2\}^{T_3 \times 1}$ represents the selection of the two weak classifiers, and $W_3 \in \mathbb{R}^{T_3 \times 1}$ contains the corresponding weights. TrainBoost is used to learn H_3 and W_3, and in this case $X = 2$.

Algorithm 2 gives a summary of our BMSR method. The overall flow is also illustrated in Fig. 1.

Algorithm 2: The proposed BMSR method

Data: training images $\{R_i : i = 1, ..., I\}$, reference dictionaries $\{D^m\}$, number of iterations T_1, T_2 and T_3, testing images $\{P_j : j = 1, ..., J\}$.
Result: learned parameters $\{H_{1,2,3}, W_{1,2,3}\}$, and labeling outputs $\{L(P_j)\}$.
Training:
$\{H_1, W_1\}$ = TrainBoost(R, SR $\{D^m\}, T_1$);
Obtain $\{Pr_1(R_i, D, n)\}$ using Eq.(3);
for $n = 1, ..., N$ **do**
  $\{H_2(:, n), W_2(:, n)\}$ = TrainBoost($R|n$, SR $\{D^m\}, T_2$);
end
Obtain $\{Pr_2(R_i, D, n)\}$ using Eq.(4);
$\{H_3, W_3\}$ = TrainBoost($R, \{Pr_1, Pr_2\}, T_3$);
Testing:
for $j = 1, ..., J$ **do**
  Obtain $Pr_1(P_j, D, n)$ using Eq.(3);
  Obtain $Pr_2(P_j, D, n)$ using Eq.(4);
  Obtain $L(P_j)$ using Eq.(5);
end

2.3. Classification of ILD Patches

For the problem of ILD classification, the aim is to classify each image patch P into one of the five categories – normal, emphysema, ground glass, fibrosis and micronodules – denoted as $L(P) = n \in \{1, ..., 5\}$.

We compute the texture-intensity-gradient features [4] as f. The patch-adaptive sparse approximation (PASA) method [4] is used in place of the basic SR. We choose PASA since it provides better classification for ILD patches [4, 10], and any SR-based classifiers can be plugged into BMSR without changing the overall method flow.

The dataset is divided into four subsets following [10]. Given a test subject, all image patches from the other subjects are collected as training images. Four dictionary subsets are then created: $\{D^m = \{D^m_n : n = 1, ..., 5\} : m = 1, ..., 4\}$. PASA classification using one dictionary subset D^m is thus one weak classifier, and four weak classifiers are used for folds 1 and 2. With such constructs of training images and weak classifiers, BMSR is then performed to learn $H_{1,2,3}$ and $W_{1,2,3}$ and classify the image patches within the test subject, as shown in Algorithm 2. We found two iterations, i.e. $T_1 = T_2 = T_3 = 2$, give good results for this dataset.

Note that the boosted training and classification are very fast, i.e. about 1 s for each test subject, since the classification outputs using the weak classifiers can be computed and stored first. The overall computational cost is thus mainly attributed to four PASA classifications for each image patch.

3. RESULTS

In this study, the publicly available ILD database [9] was used for evaluation. The database contains 113 HRCT lung images, and a total of 2062 2D ROIs are annotated with 17 different ILD tissue patterns. Following the previous study [10], we used 93 HRCT image sets to classify image patches showing the five most common tissue patterns: normal (NM), emphysema (EP), ground glass (GG), fibrosis (FB) and micronodules (MN). A total of 23131 image patches were gathered by dividing the annotated 2D ROIs into half-overlapping 31×31 pixel patches. The numbers of image patches belonging to each of the five tissue categories were 6438, 1474, 2974, 4396, 7849, respectively.

Table 1 shows the confusion matrix of the classification results using our BMSR method. There were considerable tendencies of NM misclassified as MN, EP as NM, and FB as GG. These can be explained by the high visual similarity between these tissue categories. In addition, the misclassification tended not to be reciprocal. For example, while EP was most likely to be misclassified as NM, the majority of misclassification for NM was MN. This demonstrates that EP is more similar to NM while NM is more similar to MN.

We compared the classification performance of (i) the proposed BMSR method, with the following approaches: (ii) SR-All: SR using the entire training set; (iii) SR-Subset: SR using one of the four subsets, which was the subset containing the test subject [10]; (iv) Fold 1: boosted SR with fold 1 only; and (v) Fold 2: boosted SR with fold 2 only. In all the compared

Table 1. Classification confusion matrix.

Ground Truth	Prediction				
	NM	EP	GG	FB	MN
NM	0.780	0.024	0.059	0.001	0.137
EP	0.183	0.772	0.000	0.045	0.001
GG	0.039	0.001	0.814	0.067	0.080
FB	0.016	0.040	0.100	0.824	0.021
MN	0.057	0.001	0.071	0.010	0.862

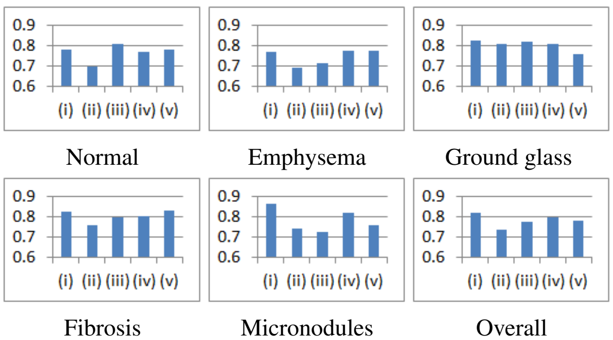

Fig. 2. Classification recall: (i) The proposed BMSR method, (ii) SR-All, (iii) SR-Subset, (iv) Fold 1, (v) Fold 2.

approaches, SR referred to the PASA classifier, and the test subject was excluded from the reference dictionary. Recall and precision were used as the performance metrics.

As shown in Fig. 2 and 3, BMSR achieved the best classification performance compared to the other approaches. We found that for many test images, the learned H_3 comprised only a certain fold, i.e. fold 1 or 2, since the classification results using that fold delivered dominant performance over the other fold; while for the others, the final classification from fold 3 was a weighted combination of the first two folds. With this boosting-based approach, BMSR thus provided improved performance over the individual folds. Folds 1 and 2 also exhibited overall advantage over SR-All and SR-Subset. This suggests that the boosted SR approach was more effective than SR based on the entire training set, and also more effective than using a single subset.

4. CONCLUSION

In this work, we present a new Boosted Multifold Sparse Representation method, to classify images in a threefold boosting manner with sparse representation on training subset as weak classifiers. The proposed method was applied to classify HRCT image patches showing five categories of ILD tissue patterns, and encouraging performance improvement was obtained. In future work, we will conduct more comprehensive evaluation, including parameter analysis for the number of subsets and iteration settings. We will also investigate

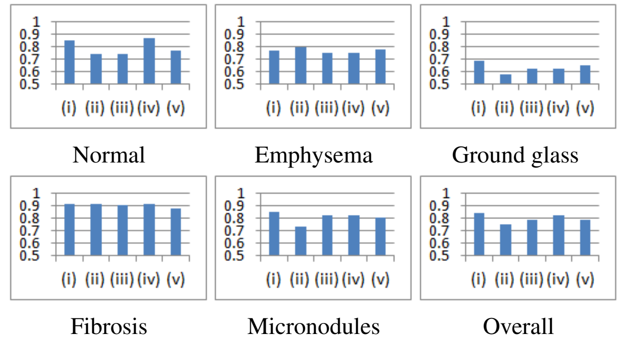

Fig. 3. Classification precision: (i) The proposed BMSR method, (ii) SR-All, (iii) SR-Subset, (iv) Fold 1, (v) Fold 2.

integrating this design with the discriminative feature learning [10] to further improve the classification performance.

5. REFERENCES

[1] M. Liu, L. Lu, X. Ye, S. Yu, and M. Salganicoff, "Sparse classification for computer aided diagnosis using learned dictionaries," *in MICCAI, LNCS*, vol. 6893, pp. 41–48, 2011.

[2] X. Huang, D. P. Dione, C. B. Compas, X. Papademetris, B. A. Lin, A. J. Sinusas, and J. S. Duncan, "A dynamical appearance model based on multiscale sparse representation: segmentation of the left ventricle from 4D echocardiography," *in MICCAI, LNCS*, vol. 7512, pp. 58–65, 2012.

[3] Y. Gao, S. Liao, and D. Shen, "Prostate segmentation by sparse representation based classification," *in MICCAI, LNCS*, vol. 7512, pp. 451–458, 2012.

[4] Y. Song, W. Cai, Y. Zhou, and D. D. Feng, "Feature-based image patch approximation for lung tissue classification," *IEEE Trans. Med. Imag.*, vol. 32, no. 4, pp. 797–808, 2013.

[5] T. Tong, R.Wolz, P. Coupe, J. V. Hajnal, D. Rueckert, and ANDI, "Segmentation of MR images via discriminative dictionary learning and sparse coding: application to hippocampus labeling," *NeuroImage*, vol. 76, pp. 11–23, 2013.

[6] Y. Song, W. Cai, H. Huang, X. Wang, S. Eberl, M.J. Fulham, and D. Feng, "Similarity guided feature labeling for lesion detection," *in MICCAI, LNCS*, 2013.

[7] F. Shi, L. Wang, G. Wu, Y. Zhang, M. Liu, J. H. Gilmore, W. Lin, and D. Shen, "Atlas construction via dictionary learning and group sparsity," *in MICCAI, LNCS*, vol. 7510, pp. 247–255, 2012.

[8] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *in CLT, LNCS*, vol. 904, pp. 23–37, 1995.

[9] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P. A. Poletti, and H. Muller, "Building a reference multimedia database for interstitial lung diseases," *Comput. Med. Imaging Graph.*, vol. 36, no. 3, pp. 227–238, 2012.

[10] Y. Song, W. Cai, S. Huh, M. Chen, T. Kanade, Y. Zhou, and D. Feng, "Discriminative data transform for image feature extraction and classification," *in MICCAI LNCS*, 2013.