Large Margin Aggregation of Local Estimates for Medical Image Classification

Yang Song[1], Weidong Cai[1], Heng Huang[2], Yun Zhou[3], David Dagan Feng[1], Mei Chen[4]

[1]BMIT Research Group, School of IT, University of Sydney, Australia
[2]Computer Science and Engineering, University of Texas at Arlington, United States
[3]Johns Hopkins University School of Medicine, United States
[4]Intel Science and Technology Center on Embedded Computing, Carnegie Mellon University, United States

Abstract. Medical images typically exhibit complex feature space distributions due to high intra-class variation and inter-class ambiguity. Monolithic classification models are often problematic. In this study, we propose a novel Large Margin Local Estimate (LMLE) method for medical image classification. In the first step, the reference images are subcategorized, and local estimates of the test image are computed based on the reference subcategories. In the second step, the local estimates are fused in a large margin model to derive the similarity level between the test image and the reference images, and the test image is classified accordingly. For evaluation, the LMLE method is applied to classify image patches of different interstitial lung disease (ILD) patterns on high-resolution computed tomography (HRCT) images. We demonstrate promising performance improvement over the state-of-the-art.

1 Introduction

Image classification can be considered the underlying task in many medical imaging problems, such as tissue segmentation, lesion detection and disease differentiation. The classification framework typically comprises feature extraction and learning-based classification. Ideally, images of the same class should have highly similar features and images of different classes should have quite dissimilar features. Well-isolated and compact clusters would form in the feature space with each cluster representing a certain class of images. In real cases, however, features of the same class could naturally be grouped into multiple clusters due to large intra-class variation, while the feature space separation between different classes could be unclear due to high inter-class ambiguity. It would thus be difficult to build a monolithic model, e.g. support vector machine (SVM) and Bayesian classifiers, to represent and classify each class accurately.

To tackle this problem, sub-categorization has recently been proposed for general imaging problems, such as face recognition [12], head orientation recognition [3], and object detection and classification [2]. By clustering the feature space of each class into multiple subcategories, locally adaptive classifiers are generated and the classification performance would be improved. These approaches normally assume the clustering results would automatically correspond to accurate classification or require only simple fusion of the subcategory results, hence the design emphases have been on the clustering techniques such as graph shift [2] and discriminative optimizations [12, 3]. However, for cases with complex feature space distributions due to intra-class variation and inter-class ambiguity, it might be a real challenge to obtain good clustering.

Alternatively, sparse representation (SR) has been widely used for medical imaging analysis in place of parametric classifiers [6, 9, 4, 8, 11, 7]. Rather than parametric modeling of the feature space separation, SR classifies a feature vector based on its reconstruction error with reference dictionaries. The classification is locally adaptive to the testing data, and could potentially accommodate complex feature space distributions. However, since the optimization goal of SR is to minimize the reconstruction error even for the wrong classes, the optimization process is generally not directly related to the classification objective and hence the classification accuracy might not be very impressive when compared to the discriminative classifiers such as SVM.

In this work, we propose a new Large Margin Local Estimate (LMLE) method to classify medical images with large intra-class variation and inter-class ambiguity. Our method consists of two main components: (1) local estimate computation - by clustering the reference dictionaries into subcategories, the subcategory-level estimates are derived for the test image with SR; and (2) large margin aggregation - the test image is classified by fusing the top-ranked local estimates in a learning-based large margin model. Our design novelties are summarized as follows. First, different from the existing sub-categorization approaches, our method of subcategory clustering is less complicated and we design a large margin aggregation model to combine the subcategory results for more accurate classification. Second, different from the standard SR that uses the entire reference dictionary, we derive local estimates based on the subcategories so that a relatively large reconstruction error could be obtained for the wrong classes. Third, while our large margin aggregation model is conceptually similar to the large margin nearest neighbor (LMNN) algorithm [10], our formulation is to minimize the distance between the test image and the sparse reconstruction from the correct class and penalize better reconstruction from the wrong classes. Lastly, we have applied our LMLE method to classify five ILD disease types on a large HRCT database and obtained promising performance improvement. The proposed method can be applicable to other classification problems as well.

2 Methods

Assume L sets of reference images corresponding to L classes are available. For each reference image, its H-length feature vector is precomputed and denoted by $r_i^l \in \mathbb{R}^H$, with $l \in \{1, ..., L\}$ as the class label, i is the index of the image in the lth reference set. The reference set of class l is represented by $R_l = \{r_i^l : i = 1, ..., N_l\}$. Given a test image I with feature vector $f \in \mathbb{R}^H$, our aim is to classify

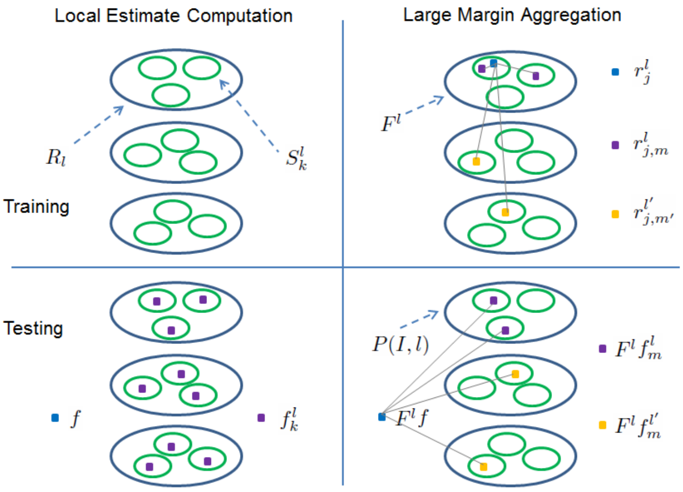

Fig. 1. Method illustration. This example assumes the dataset contains three classes, each reference set is sub-categorized into three clusters, and the classification is based on top two local estimates. Detailed explanation is referred to Sections 2.1 and 2.2.

it into one of L classes based on the reference images $\{R_l\}$. Figure 1 illustrates the overall design of the proposed LMLE method. In this section, r and f are semantically the same and both denote the image feature vectors, but r refers to the reference images specifically while f indicates the test image.

2.1 Local Estimate Computation with Sub-categorization

In the first step, we derive the local estimates for the test image based on the subcategories of the reference images. In standard SR, the underlying assumption is that a better reconstruction corresponds to higher similarity between the test image and the reference images, and the class is determined by the most similar reference set. However, with large intra-class variation, a good reconstruction for the wrong class would become highly possible by combining reference images with quite different features. We thus propose to sub-categorize the reference set of each class to minimize the feature variation within each subcategory, and the sparse reconstruction at the subcategory-level represents the local estimate of the test image by the reference images in the subcategory. We suggest that the collection of local estimates would better describe the actual similarity between the test image and the reference set. With the large margin aggregation method (Section 2.2), our method does not impose stringent requirement on the compactness and separation of the subcategories, and we thus choose to design a relatively simple sub-categorization algorithm.

Formally, given a reference set R_l, our aim is to cluster it into K subcategories of similar features. We measure the similarity between two feature vectors r_i^l and r_j^l in R_l by two criteria: (1) feature value $\|r_i^l - r_j^l\|^2$; and (2) feature separation

$\|d_i^l - d_j^l\|^2$, in which d_i^l is a $L-1$ dimensional vector with each element representing the mean Euclidean distance between r_i^l and C reference images in $R_{l' \neq l}$ that are most similar to r_i^l, and C is set to $0.1N_{l'}$. The images belonging to one subcategory would thus have similar features and similar separation from the other classes. The clustering objective is then formulated as:

$$\underset{S^l}{\operatorname{argmin}} \sum_{k=1}^{K} \sum_{r_i^l \in S_k^l} (\|r_i^l - \mu_k^l\|^2 + \|d_i^l - \theta_k^l\|^2) = \sum_{k=1}^{K} \sum_{r_i^l \in S_k^l} \left\| \begin{pmatrix} r_i^l \\ d_i^l \end{pmatrix} - \begin{pmatrix} \mu_k^l \\ \theta_k^l \end{pmatrix} \right\|^2 \quad (1)$$

where $S^l = \{S_k^l : k = 1, ..., K\}$ represents the K subcategories of reference set R_l, μ_k^l and θ_k^l represent the mean r_i^l and d_i^l in the subcategory S_k^l. This sub-categorization problem can then be solved using k-means clustering.

Next, for a test image I with feature vector f, its local estimate $f_k^l \in \mathbb{R}^H$ by a subcategory S_k^l is computed using SR. To do this, a subcategory-level reference dictionary $D_k^l \in \mathbb{R}^{H \times N_k^l}$ is first constructed by concatenating the feature vectors of reference images in S_k^l. Here N_k^l is the number of images in S_k^l. Sparse reconstruction is then formulated to derive the local estimate f_k^l:

$$\underset{x_k^l \geq 0}{\operatorname{argmin}} \|f - D_k^l x_k^l\|^2 + \alpha \|x_k^l\|_1; \qquad f_k^l = D_k^l x_k^l \quad (2)$$

We choose to impose the non-negativity constraint on the weight vector x_k^l to restrict the reconstruction performance, and the regularization parameter α is set to 0.1. The SLEP package [5] is used to solve the optimization problem.

2.2 Classification with Large Margin Aggregation

In the second step, we classify the test image I based on the $L \times K$ local estimates $\{f_k^l\}$. Many fusion algorithms can be applied, such as the max or mean pooling, i.e. selecting the most similar local estimate or computing the mean estimate from each class. A kNN approach is also possible, by selecting a number of local estimates that are closest to f from each class, and assigning I to the class with the highest similarity level between f and the selected estimates. However, due to large inter-class ambiguity, f could be similar to certain subcategories of the wrong classes, and the corresponding local estimates could be very close to f, hence affecting the kNN accuracy. To overcome this issue, we propose a large margin aggregation model to fuse the local estimates for classification. The main idea is to learn a transformation matrix in a large margin construct so that local estimates from the wrong classes would become more distant from f while those from the correct class would become closer. Similarity-based classification using such transformed vectors would then be more accurate. Our large margin aggregation method is designed based on the concept of LMNN, which has been widely popular in general computer vision as a learning-based kNN technique but has not been adapted for fusion of subcategory results.

Specifically, our aim is to learn a linear transformation matrix $F \in \mathbb{R}^{H \times H}$, so that Ff is more similar to Ff_m^l than $Ff_m^{l'}$, assuming the class label of f is

l and $l' \neq l$, and $m = 1, ..., M$ indexes the M local estimates from class l or l' that are most similar to f. With such F, we expect $\sum_{m=1}^{M} \|F(f - f_m^l)\|^2 < \sum_{m=1}^{M} \|F(f - f_m^{l'})\|^2$, and I would then be classified correctly.

To achieve this, we define the following cost function:

$$\begin{aligned}\varepsilon(F) = &\sum_{j=1}^{J} \sum_{m=1}^{M} \|F(r_j^l - r_{j,m}^l)\|^2 + \\ &\sum_{j=1}^{J} \sum_{m=1}^{M} \sum_{m'=1}^{M} [1 + \|F(r_j^l - r_{j,m}^l)\|^2 - \|F(r_j^l - r_{j,m'}^{l'})\|^2]_+\end{aligned} \tag{3}$$

in which r_j^l denotes a training sample of class l from the reference set R_l, J is the number of samples, $r_{j,m}^l$ represents the mth closest local estimate from the correct class l, and $r_{j,m'}^{l'}$ represents the m'th closest local estimate from the wrong class l'. The first term penalizes distances between the feature and the local estimates from the correct class. The second term $[z]_+ = \max(0, z)$ is the standard hinge loss, and penalizes cases where the feature is closer to the local estimates from the wrong classes than those from the correct class. By minimizing this cost function, r_j^l would thus be more similar to $\{r_{j,m}^l\}_m$ than $\{r_{j,m}^{l'}\}_m$ by a large margin, in the transformed feature space.

To minimize Eq. (3), we reformulate this optimization goal as a semidefinite programming problem:

$$\begin{aligned}\text{Minimize} &\sum_{j=1}^{J} \sum_{m=1}^{M} (r_j^l - r_{j,m}^l)^T X (r_j^l - r_{j,m}^l) + \sum_{j=1}^{J} \sum_{m=1}^{M} \sum_{m'=1}^{M} \xi_{jmm'} \\ \text{s.t. } &(r_j^l - r_{j,m'}^{l'})^T X (r_j^l - r_{j,m'}^{l'}) - (r_j^l - r_{j,m}^l)^T X (r_j^l - r_{j,m}^l) \geq 1 - \xi_{jmm'}; \\ &\xi_{jmm'} \geq 0; \quad X \succeq 0\end{aligned} \tag{4}$$

Here the matrix X is positive semidefinite and $X = F^T F$. This formulation is mathematically similar to the optimization problem in LMNN [10]. However, LMNN works by finding the nearest neighbors among a set of feature vectors, while our approach involves J feature-estimate sets $\{r_j^l, \{r_{j,m}^l\}_m, \{r_{j,m'}^{l'}\}_{m'}\}_j$ and there is no nearest-neighbor relationship among the feature vectors $\{r_j^l\}$. We have thus modified the LMNN solver to derive F.

In addition, in a multi-class setting, we choose to perform the classification in a one-versus-all manner. Specifically, by choosing samples $\{r_j^l\}$ from a single reference set R_l, the optimization problem in Eq. (4) is solved to derive a class-specific transformation matrix F^l, and a total of L matrices $\{F^l : l = 1, ..., L\}$ are learned. To classify a test image I, with each F^l, the probability of I belonging to class l is computed as:

$$P(I, l) = 1 - \frac{\sum_{m=1}^{M} \|F^l(f - f_m^l)\|^2}{\sum_{m=1}^{M} (\|F^l(f - f_m^l)\|^2 + \|F^l(f - f_m^{l'})\|^2)} \tag{5}$$

with $l' \neq l$. The class label of I then corresponds to the F^l that generates the highest probability, i.e. label$(I) = \text{argmax}_l \, P(I, l)$.

Table 1. Confusion matrix of ILD classification.

Ground Truth	Prediction (%)				
	NM	EM	GG	FB	MN
NM	**86.7**	6.1	1.6	1.1	4.5
EM	13.5	**75.0**	0.3	11.1	0.0
GG	7.6	0.0	**80.0**	7.3	5.1
FB	0.3	1.7	7.6	**86.6**	3.8
MN	3.1	0.0	5.3	4.4	**87.1**

2.3 Application to ILD Image Classification

We experimented our LMLE method on HRCT lung images from 93 ILD subjects [1]. This database is publicly available, and annotated ground truth is provided indicating 2D region-of-interest (ROI) and the associated ILD type. Similar to the state-of-the-art in this problem domain [1, 8], we performed classification on 2D image patches of 31 × 31 pixels. The dataset comprised a total of 24084 image patches of five ILD types: 6438 normal (NM), 1474 emphysema (EM), 2974 ground glass (GG), 4396 fibrosis (FB) and 7849 micronodules (MN) patches. Our objective was thus to classify the image patches into the five ILD types.

We used the texture-intensity-gradient (TIG) feature vector [8] to represent each image patch. The dataset was divided sequentially into four subsets of similar numbers of subjects. For each subset, a leave-one-subject-out scheme was applied for training and testing. 10% of this training set was then selected randomly to learn the transformation matrices $\{F^l\}$. We used only a subset in order to reduce the number of conflicting constraints and speed up the learning process. We expected that only parameters K and M needed to be tuned for different applications. In our case, the number of subcategories was set to $K = \lceil N_l/50 \rceil$ based on the size of the reference set, so that on average each subcategory would contain around 50 images. The number of top local estimates was set to $M = 5$, which was found to provide the best classification among $M = 1$ to 7.

3 Results

Table 1 shows the confusion matrix of ILD image classification using our LMLE method. We obtained more than 80% sensitivity for four of the five classes. EM was the most difficult class, with 75% sensitivity. EM patches exhibited very high visual variation, with some easily mistaken as NM while some appearing similar to FB. The number of EM patches was also small, compared to the other classes, hence there could be insufficient amount of reference images to represent the varying image features and this would affect the classification accuracy.

We compared the classification recall and precision between our proposed LMLE method and five other approaches: (1) SR – the standard SR classification without reference sub-categorization; (2) NNLE – similar to our method but replacing the large margin aggregation with kNN; (3) PASA – the patch adaptive

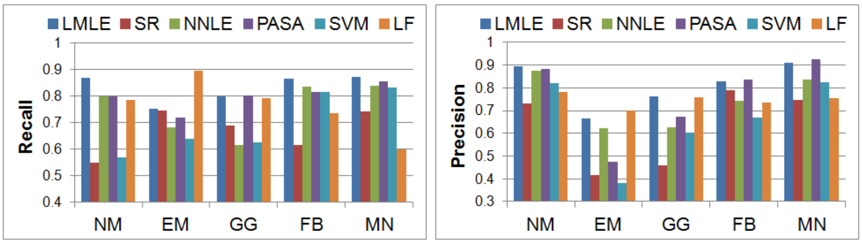

Fig. 2. The classification recall and precision, comparing our LMLE method with SR, NNLE, PASA [8], SVM and LF [1].

sparse approximation [8], which is the state-of-the-art in ILD classification and is a modified SR scheme with reference adaptation; (4) SVM – the polynomial kernel performed the best; and (5) LF – the original results using localized features published with the ILD database [1]. The compared approaches (1)–(4) used the same TIG feature as our LMLE method. Hard classification was performed without tuning the balance between precision and recall.

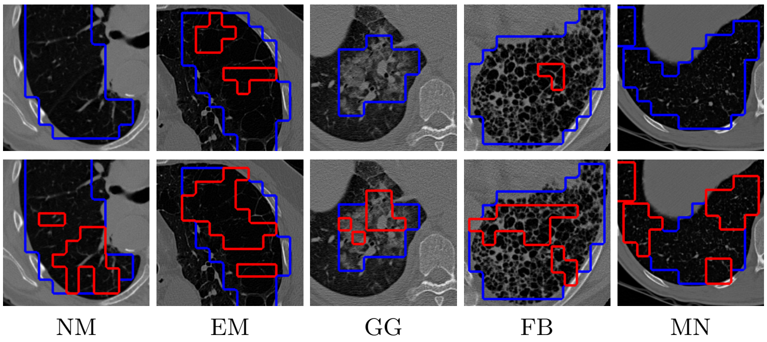

Fig. 3. Examples comparing the classification outputs between our LMLE method (top row) and SR (bottom row) for the five ILD class. The blue lines enclose the the patches that are accurately classified, while the red lines indicate the misclassified patches.

As shown in Fig. 2, our LMLE method outperformed the compared approaches. SR and SVM provided overall comparable results based on the same reference sets. The performance difference between SR and NNLE demonstrates the advantage of the reference sub-categorization. The improvement of LMLE over NNLE indicates the benefit of the large margin aggregation. While LMLE provided slightly lower precision for FB and MN than PASA, overall LMLE achieved notable gain over PASA. LMLE also demonstrated considerable improvement over LF in four of the five ILD types.

Fig. 3 illustrates some classification results, comparing our LMLE method with the SR approach. The selected images also help to show the visual similarity between different classes. Take the EM results as an example. In this case, the EM patches look indeed very similar to normal tissue, hence they tend to be misclassified as NM. By explicitly handling the intra-class variation and inter-class ambiguity, our LMLE method is effective in reducing such misclassification.

4 Conclusion

In this work, we proposed a new Large Margin Local Estimate (LMLE) method for medical image classification. The method is designed to tackle the large intra-class variation and inter-class ambiguity with two main components: local estimate computation with sub-categorization and classification with large margin aggregation. The proposed method has been applied to classify five types of interstitial lung disease (ILD) patterns on a publicly available HRCT database, and has shown consistent advantage over the compared approaches.

References

1. Depeursinge, A., Vargas, A., Platon, A., Geissbuhler, A., Poletti, P.A., Muller, H.: Building a reference multimedia database for interstitial lung diseases. Comput. Med. Imaging Graph. 36(3), 227–238 (2012)
2. Dong, J., Xia, W., Chen, Q., Feng, J., Huang, Z., Yan, S.: Subcategory-aware object classification. In: CVPR, pp. 827–834 (2013)
3. Hoai, M., Zisserman, A.: Discriminative sub-categorization. In: CVPR, pp. 1666–1673 (2013)
4. Liao, S., Gao, Y., Shen, D.: Sparse patch based prostate segmentation in CT images. In: Ayache, N. et al. (Eds) MICCAI 2012, Part III. LNCS, vol. 7512, pp. 385–392, Springer, Heidelberg (2012)
5. Liu, J., Ji, S., Ye, J.: SLEP: sparse learning with efficient projections. Arizona State University (2009), `http://www.public.asu.edu/~jye02/Software/SLEP/`
6. Liu, M., Lu, L., Ye, X., Yu, S., Salganicoff, M.: Sparse classification for computer aided diagnosis using learned dictionaries. In: Fichtinger, G. et al. (eds.) MICCAI 2011, Part III. LNCS, vol. 6893, pp. 41–48, Springer, Heidelberg (2011)
7. Song, Y., Cai, W., Huang, H., Wang, X., Zhou, Y., Fulham, M., Feng, D.: Lesion detection and characterization with context driven approximation in thoracic FDG PET-CT images of NSCLC studies. IEEE Trans. Med. Imag. 33(2), 408–421 (2014)
8. Song, Y., Cai, W., Zhou, Y., Feng, D.: Feature-based image patch approximation for lung tissue classification. IEEE Trans. Med. Imag. 32(4), 797–808 (2013)
9. Wang, H., Nie, F., Huang, H., Risacher, S., Ding, C., Saykin, A.J., Shen, L., ADNI: Sparse multi-task regression and feature selection to identify brain imaging predictors for memory performance. In ICCV pp. 557–562 (2011)
10. Weinberger, K., Saul, L.: Distance metric learning for large margin nearest neighbor classification. Journal of Machine Learning Research 10, 207–244 (2009)
11. Xu, Y., Gao, X., Lin, S., Wong, D.W.K., Liu, J., Xu, D., Cheng, C., Cheung, C.Y., Wong, T.Y.: Automatic grading of nuclear cataracts from slit-lamp lens images using group sparsity regression. In: Mori, K. et al. (eds.) MICCAI 2013, Part II. LNCS, vol. 8150, pp. 468–475, Springer, Heidelberg (2013)
12. Zhu, M., Martinez, A.: Subclass discriminant analysis. IEEE Trans. Pattern Anal. Mach. Intell. 28(8), 1274–1286 (2006)