

Discriminative Sparse Coding on Multi-Manifold for Data Representation and Classification

Jing-Yan Wang

Mathematical and Computer Sciences and Engineering Division,
King Abdullah University of Science and Technology,
Thuwal 23955-6900, Saudi Arabia
Email: jingyan.wang@kaust.edu.sa

August 21, 2012

Abstract

Sparse coding has been popularly used as an effective data representation method in various applications, such as computer vision, medical imaging and bioinformatics, etc. However, the conventional sparse coding algorithms and its manifold regularized variants (graph sparse coding and Laplacian sparse coding), learn the codebook and codes in a unsupervised manner and neglect the class information available in the training set. To address this problem, in this paper we propose a novel discriminative sparse coding method based on multi-manifold, by learning discriminative class-conditional codebooks and sparse codes from both data feature space and class labels. First, the entire training set is partitioned into multiple manifolds according to the class labels. Then, we formulate the sparse coding as a manifold-manifold matching problem and learn class-conditional codebooks and codes to maximize the manifold margins of different classes. Lastly, we present a data point-manifold matching error based strategy to classify the unlabeled data point. Experimental results on somatic mutations identification and breast tumors classification in ultrasonic images tasks demonstrate the efficacy of the proposed data representation-classification approach.

1 Introduction

Sparse coding (Sc) [1] has been successfully applied in many pattern recognition applications as a part-based data representation method, such as face recognition [2], speech recognition [3], handwritten digits recognition [4] and image clustering [4], etc. Given a set of data feature vectors organized as an input data matrix, Sc aims at finding a basis vectors pool (also known as codebook), and selecting as few basis vectors as possible from the codebook to linearly reconstruct the data feature vectors, meanwhile keeping the reconstruction error as small as possible [1].

Due to the "overcomplete" or "sufficient" characteristic of the codebook learned by Sc, the locality of the data points to be encoded might be ignored. As a result, similar data vectors may be represented as totally different sparse codes based on such codebooks, bringing the the instability of the sparse coding and hampering the robustness of the sparse coding based pattern recognition applications [5, 6]. To overcome this disadvantage, Graph regularized Sparse Coding (GraphSc) and Laplacian Sparse coding (LSc) have been proposed by Zheng et al. [4] and Gao et al. [5, 6] separately. In both these two methods, the local geometrical structure of the dataset is explicitly explored by building a k -nearest neighbor graph, and the graph Laplacian is used as a smooth operator to preserve the local manifold structure. Thus, the learned sparse coding vary smoothly along the geodesics of the data manifold [4, 5, 6].

For most pattern recognition tasks, such as somatic mutations identification [7], breast tumors classification [8], etc., the class labels are available for the training set. Using these class labels, more discriminative sparse codes are supposed to be learned in a supervised manner. However, the LSc or GraphSc are both unsupervised algorithms, thus do not utilize the class labels and ignore the discriminative information contained in the labels. Moreover, both GraphSc or LSc assume that the data points from different classes define a single general manifold in the feature space and seek common codebook and coding strategy for all data points so that the nearby points are likely to have similar codes. However, as argued by Lu et al. [9, 10], "it is still unknown that whether a single manifold could well model the data and guarantee the best recognition accuracy", thus such this assumption is arguably the most suitable.

To solve the problems mentioned above, we assume that the optimal codebooks and coding strategy for for each class should be different due to the intrinsic differences of different classes, and propose a novel supervised sparse coding method by learning discriminative codes from both the data features and class labels. We model the data points from each class as a manifold such that we can learn optimal codebook and cods for each specific class. First, we partition the entire data set into several class-conditional subsets according to the labels, and assume that each subset lay on a class-conditional manifold, which should be spanned by a independent class-conditional codebook. Instead of regularize the codes with a single manifold as in LSc and GraphSc, we apply a multi-manifold framework for sparse coding regularization. A manifold is estimated for each class. Then, we formulate the spars coding as a class-conditional data features reconstruction and manifold-manifold matching problem and learn multiple codebooks and codes to maximize the manifold margins of different classes. Lastly, we present a data point-manifold matching error based strategy to classify the unlabeled data point. Experimental results on breast tumors classification in ultrasonic images [8] and somatic mutations identification [7] tasks demonstrate the efficacy of the proposed data representation-classification approach.

2 Discriminative Sparse Coding on Multi-Manifold (DisScMM)

In this section, we will introduce the newly proposed sparse coding method on multi-manifold.

2.1 Object Function

Let us denote the training data set as $\mathcal{X} = \{x_i\} \in R^D$, $i = 1, \dots, N$, where N is the number of data points and D is the dimensionality of feature vectors of the data point, and the class labels as $\mathcal{Y} = \{y_i\} \in \mathcal{L}$, $i = 1, \dots, N$, where $\mathcal{L} = \{1, \dots, L\}$ is the set of class labels. We first divide the data set \mathcal{X} into L class-conditional subsets as $\mathcal{X}_l = \{x_i | y_i = l, x_i \in \mathcal{X}\}$, according to the class labels. Let \mathcal{X}_l be the data set of the l -th class, represented by a manifold \mathcal{M}_l . The object function of DisScMM is composed of two terms as follows.

2.1.1 Sparse Reconstruction Loss Term

Different from traditional Sc methods, we represent the data points in each class with class-conditional codebook, so that they can be better separated when the codebook and coding are selected to be different in the low-dimensional code spaces. Given a class-conditional data set \mathcal{X}_l , let $U_l = [u_{l1}, \dots, u_{lK}] \in \mathbb{R}^{D \times K}$ be the its class-conditional codebook matrix, where each $u_{lk} \in \mathbb{R}^D$ represents a code word vector in the codebook, and $v_{li} \in \mathbb{R}^K$ be the coefficient vector of $x_i \in \mathcal{X}_l$, which is the sparse coding of this data point. Each data point $x_i \in \mathcal{X}_l$ can be reconstructed as a sparse linear combination of those code word vectors in the codebook as $x_i = U_l v_{li}$. A good coding v_{li} together with codebook U_l should minimize the reconstruction loss function, and also should keep the reconstruction coefficients as sparse as possible, which can be formalized as

$$\begin{aligned} \min_{U_l, V_l} \mathcal{R}(U_l, V_l) &= \min_{U_l, V_l} \sum_{i: x_i \in \mathcal{X}_l} (\|x_i - U_l v_{li}\|^2 + \alpha \|v_{li}\|_1) \\ \text{s.t. } \|u_{lk}\|^2 &\leq c, \quad k = 1, \dots, K. \end{aligned} \quad (1)$$

where V_l is the coefficient matrix, each column of V_l is a sparse representation for a data point, and $\|v_{li}\|_1$ is a l_1 norm function to measure the sparseness of v_{li} .

2.1.2 Large Margin Term

Given a sample $x_i \in \mathcal{X}_l$ belonging to l -th class, two kinds of neighbors in the data set \mathcal{X} are considered: intra-class neighbors \mathcal{N}_i^{intra} and inter-class neighbors \mathcal{N}_i^{inter} . Intra-class neighbors of x_i are the p nearest data points from the same class as x_i , while inter-class neighbors are the the p nearest data points from different class from x_i . Using Gaussian kernel, we first define the class-conditional intra-class affinity matrix W_l^{intra} and the inter-class matrix W_l^{inter}

to characterize the similarity between $x_i \in \mathcal{X}_l$ and its neighbors in \mathcal{N}_i^{intra} as well as that between $x_i \in \mathcal{X}_l$ and \mathcal{N}_i^{inter} , respectively,

$$W_{lij}^{intra} = \begin{cases} \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}), & \text{if } x_i \in \mathcal{X}_l, \text{ and } (x_j \in \mathcal{N}_i^{intra} \text{ or } x_i \in \mathcal{N}_j^{intra}) \\ 0, & \text{otherwise} \end{cases}$$

$$W_{lij}^{inter} = \begin{cases} \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2}), & \text{if } x_i \in \mathcal{X}_l, \text{ and } (x_j \in \mathcal{N}_i^{inter} \text{ or } x_i \in \mathcal{N}_j^{inter}) \\ 0, & \text{otherwise} \end{cases} \quad (2)$$

From the viewpoint of classification, the intra-class variance should be minimized while the inter-class separability should be maximized in the sparse coding spaces, so that the class margin can be maximized for sparse coding. To this end, the large margin term of sparse coding is formulated as the following optimization problem for l -th class:

$$\begin{aligned} \min_{V_l} \mathcal{M}(V_l), \quad \mathcal{M}(V_l) = & \frac{1}{2} \sum_{i: x_i \in \mathcal{X}_l} \left(\sum_{j: x_j \in \mathcal{N}_i^{intra}} \|v_{li} - v_{lj}\|^2 W_{lij}^{intra} \right) \\ & - \frac{1}{2} \sum_{i: x_i \in \mathcal{X}_l} \left(\sum_{j: x_j \in \mathcal{N}_i^{inter}} \|v_{li} - v_{lj}\|^2 W_{lij}^{inter} \right) \end{aligned} \quad (3)$$

On the one hand, the first term of objective function of $\mathcal{M}(V_l)$ in (3) is to ensure that if x_i and x_j are close and from the same class, then their class-conditional sparse codes v_{li} and v_{lj} representations are close as well. On the other hand, the second term of objective function of $\mathcal{M}(V_l)$ in (3) ensures that if x_i and x_j are close and from different classes, then their class-conditional sparse codes v_{li} and v_{lj} representations are separated as far as possible.

2.1.3 Object Function of DisScMM

To construct the object function, we first construct the class-conditional manifold by including the intra and inter-class neighbors of data points $x_i \in \mathcal{X}_l$, as $\mathcal{M}_l = \bigcup_{i: x_i \in \mathcal{X}_l} (\{x_i\} \cup \mathcal{N}_i^{intra} \cup \mathcal{N}_i^{inter})$. The data points in this manifold of l -th class are organized as a data matrix $X_l = [x_n] \in \mathbb{R}^{D \times N_l}$, $n = 1, \dots, N_l$, $x_n \in \mathcal{M}_l$, where $N_l = |\mathcal{M}_l|$ is the number of data points in \mathcal{M}_l . The corresponding sparse coding coefficient matrix is denoted as $V_l = [v_n] \in \mathbb{R}^{K \times N_l}$, where each column v_{ln} is a sparse representation for a data point x_n . Then, with the above defined two object function terms in section 2.1.1 and section 2.1.2, we will have

the object function of DisScMM by combining them, as

$$\begin{aligned}
\mathcal{O}(U_l, V_l) &= \mathcal{R}(U_l, V_l) + \beta \mathcal{M}(V_l) \\
&= \|X_l - U_l V_l\|^2 + \alpha \sum_{n=1}^{N_l} \|v_{ln}\|_1 \\
&\quad + \beta \frac{1}{2} \sum_{n,m=1}^{N_l} \|v_{ln} - v_{lm}\|^2 W_{lnm}^{intra} - \beta \frac{1}{2} \sum_{n,m=1}^{N_l} \|v_{ln} - v_{lm}\|^2 W_{lnm}^{inter} \\
&= \|X_l - U_l V_l\|^2 + \alpha \sum_{n=1}^{N_l} \|v_{ln}\|_1 + \beta [Tr(V_l L_l^{intra} V_l^\top) - Tr(V_l L_l^{inter} V_l^\top)] \\
&= \|X_l - U_l V_l\|^2 + \alpha \sum_{n=1}^{N_l} \|v_{ln}\|_1 + \beta Tr(V_l L_l V_l^\top)
\end{aligned} \tag{4}$$

where $L_l^{intra} = D_l^{intra} - W_l^{intra}$ and $L_l^{inter} = D_l^{inter} - W_l^{inter}$ are the Laplacian matrices, D_l^{intra} and D_l^{inter} are diagonal matrix whose entries are $D_{l_{nn}}^{intra} = \sum_{m=1}^{N_l} W_{lnm}^{intra}$ and $D_{l_{nn}}^{inter} = \sum_{m=1}^{N_l} W_{lnm}^{inter}$ separately, $L_l = L_l^{intra} - L_l^{inter}$, where β is the trade-off parameter.

With the defined object function, we formulate the proposed DisScMM as the following optimization problem:

$$\begin{aligned}
&\min_{U_l, V_l} \mathcal{O}(U_l, V_l) \\
&s.t. \|u_{lk}\|^2 \leq c, \quad k = 1, \dots, K.
\end{aligned} \tag{5}$$

Note that for each manifold, such a optimization will be performed to learn a class-conditional codebook and the codes.

2.2 Optimization

The optimal U_l and V_l of (5) can be solved by following the iteratively optimization method introduced in GraphSc [4] or LapSc [5, 6]. An iterative, two-step strategy is adopted to alternately optimize U_l and V_l . At each iteration, one of U_l and V_l is optimized while the other is fixed, and then the roles of U_l and V_l are switched. Iterations are repeated until a maximum number of iterations is reached.

2.2.1 On Optimizing Codebooks U_l

By fixing V_l the optimization problem (5) is reduced to

$$\begin{aligned}
&\min_{U_l} \|X_l - U_l V_l\|^2 \\
&s.t. \|u_{lk}\|^2 \leq c, \quad k = 1, \dots, K.
\end{aligned} \tag{6}$$

The solution of this problem is introduced in [1] as

$$U_l^* = X_l V_l^\top (V_l V_l^\top + \text{diag}(\lambda^*))^{-1} \tag{7}$$

where $\lambda = [\lambda_1, \dots, \lambda_K]^\top$, λ_k is the Lagrange multiplier associated with the k -th inequality constraint $\|u_{lk}\|^2 \leq c$, and λ^* is the optimal solution of λ . For more details, we refer the readers to [1, 4].

2.2.2 On Optimizing Sparse Codes V_l

By fixing U_l , the optimization problem (5) becomes

$$\min_{V_l} \|X_l - U_l V_l\|^2 + \alpha \sum_{n=1}^{N_l} \|v_{ln}\|_1 + \beta \text{Tr}(V_l L_l V_l^\top) \quad (8)$$

Each coding vector v_{ln} is optimized one by one. To optimize v_{ln} , we fix all the remaining sparse codes $v_{lm} (m \neq n)$. Note that the Laplacian regularizer of multi-manifold can be rewritten as $\text{Tr}(V_l L_l V_l^\top) = \sum_{n,m=1}^{N_l} L_{nm} v_{ln}^\top v_{lm}$. Then (8) is further reduced to

$$\min_{v_{ln}} \|x_n - U_l v_{ln}\|^2 + \alpha \|v_{ln}\|_1 + \beta \left[L_{nn} v_{ln}^\top v_{ln} + 2v_{ln}^\top \sum_{m \neq n} L_{nm} v_{lm} \right] \quad (9)$$

This problem can be optimized by the graph regularized Sparse Codes learning introduced in Algorithm 1 of [4]. or the feature-sign search algorithm introduced in Algorithm 1 of [6]. Here we adopt the one introduced in [4]. In fact, these two algorithms are basically the same except the initialization procedure. Moreover, graph regularized Sparse Codes learning introduced in Algorithm 1 of [4] requires the graph weight matrix to be symmetric while the other one do not.

The learning procedure of DisScMM algorithm is summarized in Algorithm 1.

2.3 Classifier of DisScMM

Differently from traditional Sc methods which can only be used to represent the data, the DisScMM can also makes use of the discriminative nature of sparse coding on multi-manifold to perform classification. When a new data point x_t comes in, we match it to all the manifolds and then assign it to the class with minimum matching error. Assuming x_t belongs to l -th class, we first calculate its intra-class nearest neighbors \mathcal{N}_{lt}^{intra} and inter-class nearest neighbors \mathcal{N}_{lt}^{inter} from \mathcal{M}_l . We also suppose that the input of this new data point has no effect on the discriminate graphs in the sparse codes of \mathcal{M}_l , so the sparse codes v_{ln} for $x_n \in \mathcal{M}_l$ are fixed. Then the match error between x_t and \mathcal{M}_l is defined as:

$$\begin{aligned} \mathcal{E}_l(x_t) = \min_{v_{lt}} \|x_t - U_l v_{lt}\|^2 + \alpha \|v_{lt}\|_1 + \frac{\beta}{2} \sum_{n: x_n \in \mathcal{N}_{lt}^{intra}} \|v_{lt} - v_{ln}\|^2 W_{ltn}^{intra} \\ - \frac{\beta}{2} \sum_{n: x_n \in \mathcal{N}_{lt}^{inter}} \|v_{lt} - v_{ln}\|^2 W_{ltn}^{inter} \end{aligned} \quad (10)$$

Algorithm 1 The learning procedure of DisScMM Algorithm.

INPUT: Training sets $\mathcal{M}_1, \dots, \mathcal{M}_L$ of L classes of multi-manifold;
for $l = 1, \dots, L$ **do**
 Construct discriminate graph weight matrices as in (2) and corresponding Laplacian matrices L_l for l -th manifold.
 Initialize the class-conditional codebook U_l^0 and sparse codes V_l^0 for l -th manifold, by performing Sc to M_l .
 for $t = 1, \dots, T$ **do**
 for $n = 1, \dots, N_l$ **do**
 Update the sparse codes v_{ln}^t while fixing $v_{lm}^{t-1}, m \neq n$ and U_l^{t-1} by solving (9) for l -th manifold.
 end for
 Update the codebook U_l^t while fixing V_l^t by (7) for l -th manifold.
 end for
end for
OUTPUT: The final class-conditional codebooks U_l^T and sparse codes V_l^T , $l = 1, \dots, L$.

where W_{ltn}^{intra} and W_{ltn}^{inter} are the intra and inter-similarities of x_t to the n -th data point of \mathcal{M}_l , which is calculated by (2). This optimization problem can also be solve by Algorithm proposed in [4]. We finally assign a label y_t to x_t as follows:

$$y_t \leftarrow l^* = \underset{l \in \mathcal{L}}{\operatorname{argmin}} \mathcal{E}_l(x_t) \quad (11)$$

The classification procedure is summarized in Algorithm 2.

Algorithm 2 The classification procedure of DisScMM Algorithm.

INPUT: Training sets $\mathcal{M}_1, \dots, \mathcal{M}_L$ of L classes of multi-manifold;
INPUT: The class-conditional codebooks U_l and sparse codes V_l for L manifolds, $l = 1, \dots, L$.
INPUT: The input unlabeled data point x_t .
for $l = 1, \dots, L$ **do**
 Extend the discriminate graph weight matrices by adding x_t as in (2) and compute corresponding Laplacian matrices L_l for l -th manifold.
 Compute the matching error $\mathcal{E}_l(x_t)$ of x_t to \mathcal{M}_l as in (10).
end for
Classify x_t into the l^* -th class with minimum matching error as in (11).
OUTPUT: The class label y_t of x_t .

3 Experiments

In this section, we will evaluate the proposed method on two challenging data classification tasks.

3.1 Experiment I: Identifying Somatic Mutations

Profiling tumours for single nucleotide variant (SNV) somatic mutations using next-generation sequencing technology (NGS) plays an important role in the study of cancer genomes [7]. In this experiment, we will evaluate our DisScMM on the task of inferring somatic mutations from paired tumour/normal NGS data.

3.1.1 Database and Setup

Two independent datasets are used to train and test the performance of the DisScMM method for somatic mutation identification.

Training Set The training dataset is selected from the exome capture data containing 3369 variants which are predicted by using only allelic counts and liberal thresholds [7]. Further re-sequencing experiments revalidated 1015 somatic mutations, 471 germline and 1883 wild-type positions. Our selected training data set contains 800 somatic mutations, and 1800 non-somatic mutations (germline and wildtype).

Test Set The test dataset is selected from the whole genome shotgun data containing 113 somatic mutations, 57 germline mutations and 337 wildtypes [7]. These positions are deliberately held out of the training data so that the test set and the training set are completely independent from each other. We select 90 somatic mutations and 300 non-somatic mutations to construct the test set.

Given the i -th candidate mutation site of the genome in the dataset, it is represented by a feature vector x_i with 106 feature components constructed from both the tumor and normal data as in [7]. The somatic mutations identifying problem is to predict the label y_i of the feature represented site. y_i is defined as

$$y_i = \begin{cases} 1, & \text{if } i\text{-th site is a somatic mutation,} \\ 2, & \text{if } i\text{-th site is a non-somatic mutation.} \end{cases} \quad (12)$$

To predicate the class labels in the test set, we first learn the codebooks for the somatic mutations manifold and non-somatic mutations manifold using the training set for DisScMM. For the learning procedure, we applied a *10-fold cross-validation* analysis to find the optimal hyper-parameters. Then the learned DisScMM model will be applied to the independent test set to classify each candidate mutation site into somatic mutations or non-somatic mutations. Some competing algorithms, including Sc [1], GraphSc [4] and LapSc [6] are also tested as mutation representation methods.

To evaluate the performances of the classification results, we employ the recall, precision [11], accuracy, F-score, matthews correlation coefficient (MCC) as metrics. Recall, precision and accuracy are defined as

$$Recall = \frac{TP}{TP + FN}, Precision = \frac{TP}{TP + FP}, Accuracy = \frac{TP + TN}{TP + FN + TP + FP} \quad (13)$$

where TP, FP, TN, FN are the number of true positives, false positives, false positives and false negative respectively. The F-score is the harmonic mean of precision defined as

$$Recall = 2 \times \frac{Recall \times Precision}{Recall + Precision} \quad (14)$$

Recall Precision, accuracy, the F-score are comprised between 0 and 1, and the classifier with the larger value has the better the performance. The MCC is given by

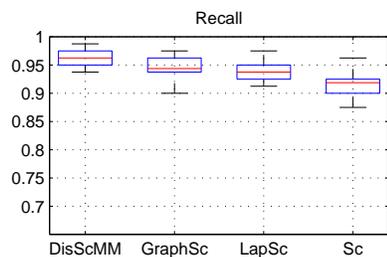
$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FN)(TP + FP)(TN + FN)(TN + FP)}} \quad (15)$$

The MCC value is between -1 and 1. A perfect classifier has $MCC = 1$, a random predictor has $MCC = 0$, while perfect inverted predictor has $MCC = -1$.

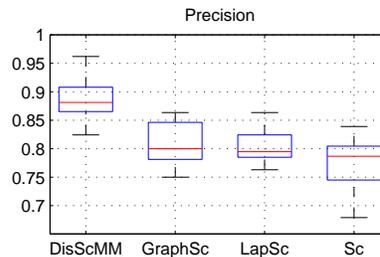
3.1.2 Results

The boxplots of recalls, precisions, accuracies, F-scores and MCCs of 10-fold cross-validation on the training data set are shown in Fig. 1 (a) - (e), respectively, where the various performance metric values of our DisScMM show accuracy of the returned top results. We observe that for all performances measures, DisScMM outperforms the baseline methods significantly in terms of both median value and Q values. We also observed that the unsupervised single general graph based sparse coding, i.e. GraphSc and Lap, has comparable performance to the each other. From these figures, it is not very surprising to see that original Sc provides the worst performance since the Sc function ignores locality of the data points.

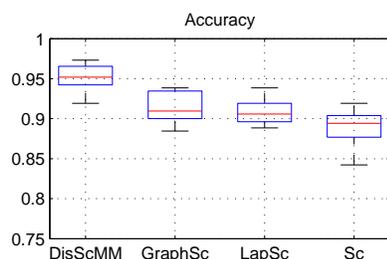
Fig. 2 summarizes the recalls, precisions, accuracies, F-scores and MMCs for the proposed DisScMM and its competitors on the independent test dataset. According to Fig. 2, we first observe a significant difference between recall and precision scores for all the methods, which is consistent with the observations reported in the previous 10-fold cross-validation on the training dataset. The possible reason is the significant unbalanced number of the positive and negative samples. Second, we observe that for all the cases, DisScMM outperforms GraphSc, LapSc and Sc significantly. Fig. 2 also shows that the sparse coding methods with manifold regularization outperform sparse coding without it. Our DisScMM based somatic mutations identifying method outperforms both



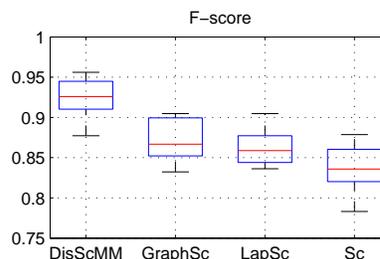
(a) Recall



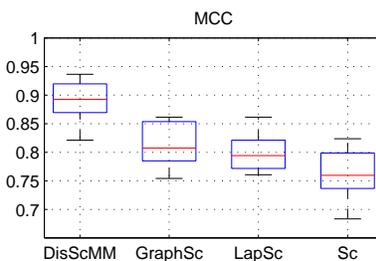
(b) Precision



(c) Accuracy



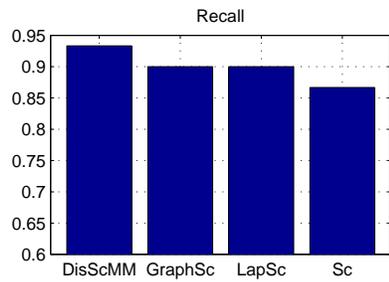
(d) F-score



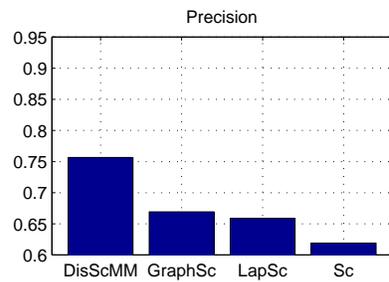
(e) MCC

Figure 1: Boxplots of recalls, precisions, accuracies, F-scores and MMCs of 10-fold cross-validation on training set of somatic mutation identification.

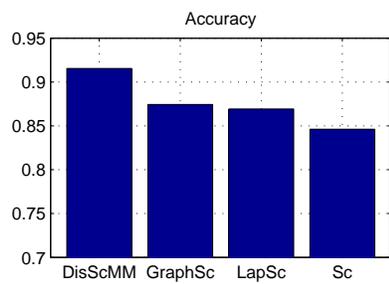
GraphSc and LapSc based tagging methods, and achieve the best somatic mutations identifying performance of all the methods, which proves the effectiveness of DisScMM for this task. Moreover, GraphSc and LapSc achieves much better



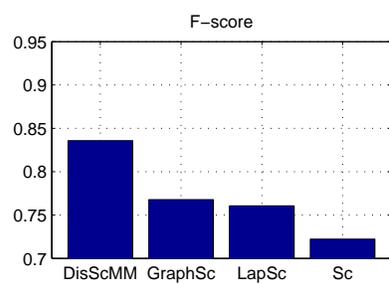
(a) Recall



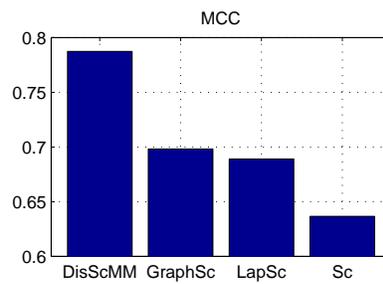
(b) Precision



(c) Accuracy



(d) F-score



(e) MCC

Figure 2: The recalls, precisions, accuracies, F-scores and MMCs on the test set of somatic mutation identification.

performance than original Sc, which proves the usefulness of regularizing the sparse code with the nearest graphs.

3.2 Experiment II: Breast Tumor Classification in Ultrasonic Images

Medical examination based on ultrasound imaging is indispensable for the early detection and treatment of breast cancers [8]. Thus, Developing automated differential diagnosis system that classifies a given breast tumor as benign or malignant plays an important role in modern medical examination. In this experiment, we will evaluate the performance in the task of breast tumor classification in ultrasonic images.

3.2.1 Dataset and Setup

We collect 340 ultrasound images for the evaluation of proposed tumor classification methods. Each of the ultrasound image included a biopsy-proven tumor (a carcinoma, a fibroadenoma, or a cyst), where carcinoma is malignant tumor while fibroadenoma and cyst are benign tumors. The tumor border is delineated manually. The data set contains 220 carcinomas, 60 fibroadenomas, and 60 cysts.

Given an ultrasound image, we extract 208 features and present them in a feature vector x . The 208 features consist of the K-related and conventional features, covering all of the diagnostic observations [8]. The classification problem is to differentiate three types of lesions (carcinoma, fibroadenoma, and cyst). For validation, we conducted a 5-fold cross-validation test. The data set is firstly divided randomly into 5 subsets and then 4 subsets were used for training, and the remaining 1 subset was used to test the proposed DisScMM, the GraphSc, LapSc, and Sc methods. We repeat the cross-validation process 5 times, and each of the 5 subsamples used exactly once as validation data.

3.2.2 Results

Fig. 3 shows the boxplots of classification accuracies obtained by different methods on the ultrasonic breast tumor images dataset. As shown in Fig. 3, our DisScMM method can achieve much better results on the 5-fold cross-validation protocol than the state-of-the-art sparse coding methods. Specifically, DisScMM outperforms almost all of the compared sparse coding methods across different tumor classes. There are two possible reasons to explain why our DisScMM method is superior to these methods:

1. our supervised method explores the discriminative information explicitly by multi-manifold regularization, while most state-of-the-art sparse coding methods are intrinsically unsupervised methods even they can extract some discriminative information from the graph model;
2. our method codes the features in a supervised manner by using class-conditional codebook and multi-manifold regularizer while others code features in an unsupervised general way.

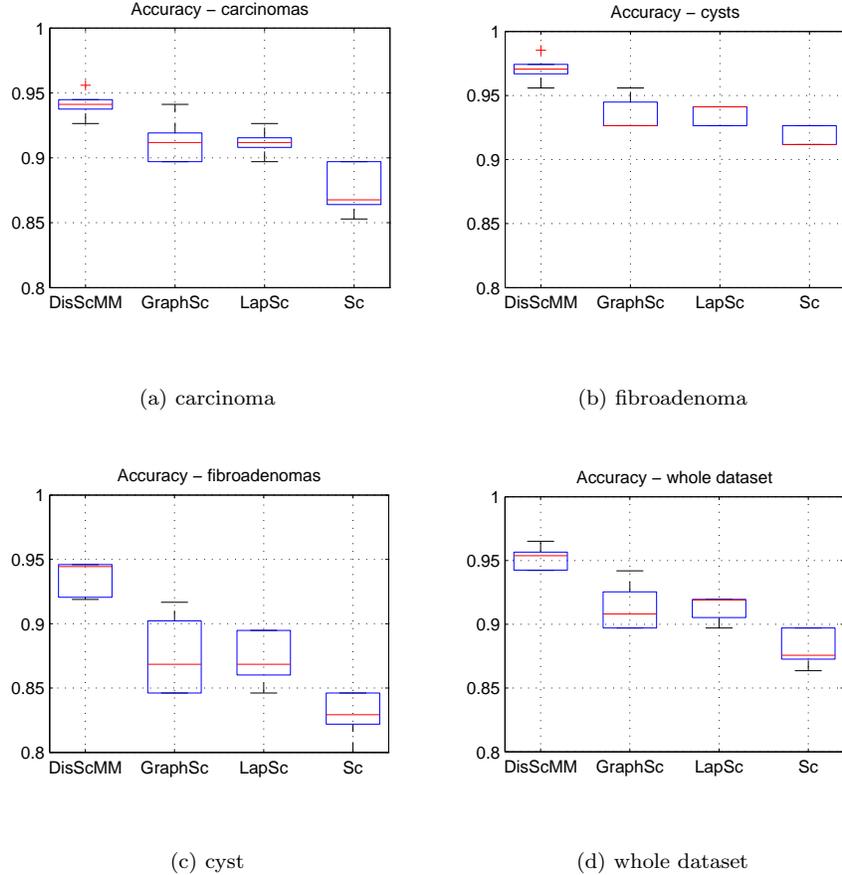


Figure 3: Boxplots of accuracies of different tumors on the ultrasonic breast tumor images set.

4 Conclusion and Future Work

In this paper we have proposed a novel discriminative sparse coding method to address the data representation and classification problem. Multiple manifolds are constructed for sub-sets of different classes. The class-conditional sparse coding are conducted to maximize the manifold margins of different classes. Experimental results on two challenging tasks are presented to demonstrate the efficacy of the proposed approach.

In the future, we are interested in designing multi-multiple regularized non-negative matrix factorization (NMF) [12] by exploring the class label to improve the data representation. Moreover, how to utilizing the coding results to further refine the manifolds model appears to be another interesting di-

rection of future work. Moreover, DisScMM can also be used to bioinformatics [13, 14, 15, 16, 17, 18], medical imaging [19, 20, 21, 22], biometrics [23, 24, 25, 26, 27, 28, 29, 30, 31, 32] and computer vision [33, 34, 35, 36].

References

- [1] H. Lee, A. Battle, R. Raina, A. Y. Ng, Efficient sparse coding algorithms, in: In NIPS, NIPS, 2007, pp. 801–808.
- [2] J. Sun, Q. Zhuo, C. Ma, W. Wang, Sparse image coding with clustering property and its application to face recognition, *Pattern Recognition* 34 (2001) 1883–4.
- [3] G. Sivaram, S. Nemala, M. Elhilali, T. Tran, H. Hermansky, Sparse coding for speech recognition, in: *Proceedings 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2010, 2010*, pp. 4346–9, 2010 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2010, 14-19 March 2010, Dallas, TX, USA.
- [4] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, D. Cai, Graph Regularized Sparse Coding for Image Representation, *IEEE TRANSACTIONS ON IMAGE PROCESSING* 20 (5) (2011) 1327–1336. doi:{10.1109/TIP.2010.2090535}.
- [5] S. Gao, I. Tsang, L.-T. Chia, P. Zhao, Local features are not lonely — laplacian sparse coding for image classification, in: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010*, pp. 3555–3561. doi:10.1109/CVPR.2010.5539943.
- [6] S. Gao, I. Tsang, L. Chia, Laplacian sparse coding, hypergraph laplacian sparse coding, and applications, *Pattern Analysis and Machine Intelligence, IEEE Transactions on PP* (99) (2012) 1. doi:10.1109/TPAMI.2012.63.
- [7] J. Ding, A. Bashashati, A. Roth, A. Oloumi, K. Tse, T. Zeng, G. Haf-fari, M. Hirst, M. A. Marra, A. Condon, S. Aparicio, S. P. Shah, Feature-based classifiers for somatic mutation detection in tumour-normal paired sequencing data, *BIOINFORMATICS* 28 (2) (2012) 167–175. doi:{10.1093/bioinformatics/btr629}.
- [8] A. Takemura, A. Shimizu, K. Hamamoto, Discrimination of Breast Tumors in Ultrasonic Images by Classifier Ensemble Trained with AdaBoost, *ELECTRONICS AND COMMUNICATIONS IN JAPAN* 94 (9) (2011) 18–29. doi:{10.1002/ecj.10356}.
- [9] J. Lu, Y. Tan, G. Wang, Discriminative multi-manifold analysis for face recognition from a single training sample per person, *Pattern Analysis and Machine Intelligence, IEEE Transactions on PP* (99) (2012) 1. doi:10.1109/TPAMI.2012.70.

- [10] J. Lu, Y.-P. Tan, G. Wang, Discriminative multi-manifold analysis for face recognition from a single training sample per person, in: 2011 IEEE International Conference on Computer Vision (ICCV 2011), 2011, pp. 1943–50, 2011 IEEE International Conference on Computer Vision (ICCV 2011), 6-13 Nov. 2011, Barcelona, Spain.
- [11] T. Zhang, D. Xu, J. Chen, Application-oriented purely semantic precision and recall for ontology mapping evaluation, KNOWLEDGE-BASED SYSTEMS 21 (8) (2008) 794–799. doi:{10.1016/j.knosys.2008.03.060}.
- [12] J.-Y. Wang, I. K. Almasri, X. Gao, Adaptive graph regularized nonnegative matrix factorization via feature selection, in: The 21st International Conference on Pattern Recognition (ICPR2012), 2012.
- [13] J. Wang, Y. Li, SEQUENTIAL LINEAR NEIGHBORHOOD PROPAGATION FOR SEMI-SUPERVISED PROTEIN FUNCTION PREDICTION, JOURNAL OF BIOINFORMATICS AND COMPUTATIONAL BIOLOGY 9 (6) (2011) 663–679. doi:{10.1142/S0219720011005550}.
- [14] Jingyan Wang, Yongping Li, Ying Zhang, Ning Tang, Chao Wang, Class Conditional Distance Metric for 3D Protein Structure Classification, in: 2011 5th International Conference on Bioinformatics and Biomedical Engineering, 2011, p. 4 pp. doi:10.1109/icbbe.2011.5780014.
- [15] Jingyan Wang, Yongping Li, Ying Zhang, Jianhua He, Semi-supervised Protein Function Prediction via Sequential Linear Neighborhood Propagation, in: Bio-Inspired Computing and Applications. 7th International Conference on Intelligent Computing, ICIC 2011.Revised Selected Papers, 2011, pp. 435–41.
- [16] J. Wang, X. Gao, Q. Wang, Y. Li, ProDis-ContSHC: learning protein dissimilarity measures and hierarchical context coherently for protein-protein comparison in protein database retrieval, BMC BIOINFORMATICS 13 (7). doi:{10.1186/1471-2105-13-S7-S2}.
- [17] J. Wang, Y. Li, Q. Wang, X. You, J. Man, C. Wang, X. Gao, ProClusEnsem: Predicting membrane protein types by fusing different modes of pseudo amino acid composition, COMPUTERS IN BIOLOGY AND MEDICINE 42 (5) (2012) 564–574. doi:{10.1016/j.combiomed.2012.01.012}.
- [18] H. Pei, J. Li, M. Lv, J. Wang, J. Gao, J. Lu, Y. Li, Q. Huang, J. Hu, C. Fan, A graphene-based sensor array for high-precision and adaptive target identification with ensemble aptamers, Journal of the American Chemical Societydoi:10.1021/ja305814u.
- [19] Jingyan Wang, Yongping Li, Ying Zhang, Honglan Xie, Chao Wang, Boosted learning of visual word weighting factors for bag-of-features

- based medical image retrieval, in: Proceedings of the Sixth International Conference on Image and Graphics (ICIG 2011), 2011, pp. 1035–40. doi:10.1109/ICIG.2011.193.
- [20] Jingyan Wang, Yongping Li, Ying Zhang, Honglan Xie, Chao Wang, Bag-of-Features Based Classification of Breast Parenchymal Tissue in the Mammogram via Jointly Selecting and Weighting Visual Words, in: Proceedings of the Sixth International Conference on Image and Graphics (ICIG 2011), 2011, pp. 622–7. doi:10.1109/ICIG.2011.192.
- [21] J. Wang, Y. Li, E. Marchiori, C. Wang, Iterated large-margin discriminant analysis for feature dimensionality reduction in medical image retrieval, International Journal of Biomedical Engineering and Technology 7 (2) (2011) 116 – 134. doi:10.1504/IJBET.2011.043174.
- [22] J. Wang, I. Almasri, Modeling multiple visual words assignment for bag-of-features based medical image retrieval, in: Proceedings of the IASTED International Conference on Computer Graphics and Imaging, CGIM 2012, Crete, Greece, 2012, pp. 217 – 224. doi:10.2316/P.2012.779-015.
- [23] J. Wang, Y. Li, P. Liang, G. Zhang, X. Ao, An Effective Multi-Biometrics Solution for Embedded Device, in: 2009 IEEE INTERNATIONAL CONFERENCE ON SYSTEMS, MAN AND CYBERNETICS (SMC 2009), VOLS 1-9, IEEE International Conference on Systems Man and Cybernetics Conference Proceedings, 2009, pp. 917–922. doi:{10.1109/ICSMC.2009.5346745}.
- [24] J. Wang, Y. Li, X. Ao, C. Wang, J. Zhou, MULTI-MODAL BIOMETRIC AUTHENTICATION FUSING IRIS AND PALMPRINT BASED ON GMM, in: 2009 IEEE/SP 15TH WORKSHOP ON STATISTICAL SIGNAL PROCESSING, VOLS 1 AND 2, 2009, pp. 349–352. doi:{10.1109/SSP.2009.5278568}.
- [25] Jingyan Wang, Yongping Li, Chao Wang, How to handle missing data in robust multi-biometrics verification, International Journal of Biometrics 3 (2011) 265–83. doi:10.1504/IJBM.2011.040819.
- [26] X. Zhai, Y. Zhao, J. Wang, Y. Li, Q. Wang, H. Xie, Adaptive SVM Fusion for Robust Multi-Biometrics Verification with Missing Data, in: 2011 INTERNATIONAL CONFERENCE ON ENERGY AND ENVIRONMENTAL SCIENCE-ICEES 2011, Vol. 11 of Energy Procedia, 2011, pp. 1006–1012. doi:{10.1016/j.egypro.2011.10.349}.
- [27] J. Liu, Y. Hou, J. Wang, Y. Li, Q. Wang, J. Man, H. Xie, J. He, Fusing Iris and Palmprint at Image Level for Multi-Biometrics Verification, in: FOURTH INTERNATIONAL CONFERENCE ON MACHINE VISION (ICMV 2011): COMPUTER VISION AND IMAGE ANALYSIS: PATTERN RECOGNITION AND BASIC TECHNOLOGIES, Vol. 8350 of Proceedings of SPIE, 2012. doi:{10.1117/12.920534}.

- [28] J. Wang, Y. Li, Y. Zhang, Y. Huang, Implementing multimodal biometric solutions in embedded systems, *Biometrics - Unique and Diverse Applications in Nature* doi:10.5772/15843.
- [29] Chao Wang, Xubo Song, Tracking facial feature points with prediction-assisted view-based active shape model, in: *Proceedings 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG 2011)*, IEEE, 2011, pp. 259–64, 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG 2011), 21-24 March 2011, Santa Barbara, CA, USA.
- [30] C. Wang, Y. Li, C. Wang, An Efficient Illumination Compensation based on Plane-fit for Face Recognition, in: *2008 10TH INTERNATIONAL CONFERENCE ON CONTROL AUTOMATION ROBOTICS & VISION: ICARV 2008, VOLS 1-4*, IEEE, 2008, pp. 939–943, 10th International Conference on Control, Automation, Robotics and Vision, Hanoi, VIETNAM, DEC 17-20, 2008.
- [31] C. Wang, W. Jiang, X. Dong, Characterization of clustered microcalcifications in mammograms based on support vector machines with genetic algorithms, in: *Proceedings of International Symposium on Biophotonics, Nanophotonics and Metamaterials*, Zhejiang Univ; Royal Inst Technol; Chinese Univ Hong Kong; SUNY Buffalo; IEEE LEOS, 2006, pp. 106–109, International Symposium on Biophotonics, Nanophotonics and Metamaterials, Hangzhou, PEOPLES R CHINA, OCT 16-18, 2006.
- [32] C. Wang, Y. Li, Combine image quality fusion and illumination compensation for video-based face recognition, *NEUROCOMPUTING* 73 (7-9, SI) (2010) 1478–1490. doi:{10.1016/j.neucom.2009.11.010}.
- [33] J. Wang, Y. Li, X. Bai, Y. Zhang, C. Wang, N. Tang, Learning context-sensitive similarity by shortest path propagation, *PATTERN RECOGNITION* 44 (10-11, SI) (2011) 2367–2374. doi:{10.1016/j.patcog.2011.02.007}.
- [34] Zhonghua Liu, Jingyan Wang, Yongping Li, Ying Zhang, Chao Wang, Quantized image patches co-occurrence matrix: a new statistical approach for texture classification using image patch exemplars, in: *Proceedings of the SPIE - The International Society for Optical Engineering*, Vol. 8009, 2011, p. 80092P (5 pp.). doi:10.1117/12.896155.
- [35] J. Wang, M. A. Jabbar, Multiple kernel learning for adaptive graph regularized nonnegative matrix factorization, in: *Proceedings of the IASTED International Conference on Signal Processing, Pattern Recognition and Applications, SPPRA 2012*, Crete, Greece, 2012, pp. 115 – 122. doi:10.2316/P.2012.778–049.
- [36] Z. Liu, J. Wang, J. Man, Y. Li, X. You, C. Wang, Self-adaptive Local Fisher Discriminant Analysis for semi-supervised image recognition, *International Journal of Biometrics*.